
Springer Handbook of Electronic and Photonic Materials

Safa Kasap, Peter Capper (Eds.)

With CD-ROM, 930 Figures and 168 Tables



Springer

Springer Handbook of Electronic and Photonic Materials

Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical sciences and engineering. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of physical and technical information.

The volumes are designed to be useful as readable desk reference books to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook of Electronic and Photonic Materials

Safa Kasap, Peter Capper (Eds.)

With CD-ROM, 930 Figures and 168 Tables



Springer

Editors:
Safa Kasap
University of Saskatchewan
Department of Electrical Engineering
Saskatoon, SK, S7N 5A9
Canada

Peter Capper
SELEX Sensors and Airborne Systems Infrared Ltd.
Millbrook Industrial Estate
Southampton, Hampshire SO15 0EG
United Kingdom

Assistant Editor:
Cyril Koughia
University of Saskatchewan
Canada

Library of Congress Control Number: 2006921596

ISBN-10: 0-387-26059-5 e-ISBN: 0-387-29185-7
ISBN-13: 978-0-387-26059-4 Printed on acid free paper

© 2006, Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

The use of designations, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Product liability: The publisher cannot guarantee the accuracy of any information about dosage and application contained in this book. In every individual case the user must check such information by consulting the relevant literature.

Production and typesetting: LE-TeX GbR, Leipzig
Handbook coordinator: Dr. W. Skolaut, Heidelberg
Typography, layout and illustrations:
schreiberVIS, Seeheim & Hippmann GbR, Schwarzenbruck
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: WMXDesign GmbH, Heidelberg
Printing and binding: Stürtz GmbH, Würzburg

SPIN 11051855 9065/3141/YL 5 4 3 2 1 0

Foreword

The Editors, Authors, and Publisher are to be congratulated on this distinguished volume, which will be an invaluable source of information to all workers in the area of electronic and photonic materials. Having made contributions to earlier handbooks, I am well aware of the considerable, and sustained work that is necessary to produce a volume of this kind. This particular handbook, however, is distinguished by its breadth of coverage in the field, and the way in which it discusses the very latest developments. In such a rapidly moving field, this is a considerable challenge, and it has been met admirably.

Previous handbooks and encyclopaedia have tended to concentrate on semiconducting materials, for the understandable reason of their dominance in the electronics field, and the wide range of semiconducting materials and phenomena that must be covered. Few have been courageous enough to predict future trends, but in 1992 Mahajan and Kimerling attempted this in the Introduction to their Concise Encyclopaedia of Semiconducting Materials and Related Technologies (Pergamon), and foresaw future challenges in the areas of nanoelectronics, low dislocation-density III-V substrates, semi-insulating III-V substrates, patterned epitaxy of III-Vs, alternative dielectrics and contacts for silicon technology, and developments in ion-implantation and diffusion. To a greater or lesser extent, all of these have been proved to be true, but it illustrates how difficult it is to make such a prediction.

Not many people would have thought, a decade ago, that the III-nitrides would occupy an important position in this book. As high melting point materials, with the associated growth problems, they were not high on the list of favourites for light emitters at the blue end of the spectrum! The story is a fascinating one – at least as interesting as the solution to the problem of the short working life of early solid-state lasers at the red end of the spectrum. Optoelectronics and photonics, in general, have seen one of the most spectacular advances over the last decade, and this is fully reflected in the book, ranging from visible light emitters, to infra-red materials. The book covers a wide range of work in Part D, including III-V and II-VI optoelectronic materials and band-gap engineering, as well as photonic glasses, liquid crystals, organic

photoconductors, and the new area of photonic crystals. The whole Part reflects materials for light generation, processing, transmission and detection – all the essential elements for using light instead of electrons.

In the Materials for Electronics part (Part C) the book charts the progress in silicon – overwhelmingly the dominant material for a whole range of electronic functions and circuitry – including new dielectrics and other issues associated with shrinking geometry of circuits and devices to produce ever higher packing densities. It also includes areas rarely covered in other books – thick films, high-temperature electronic materials, amorphous and microcrystalline materials. The existing developments that extend the life of silicon technology, including silicon/germanium alloys, appear too, and raise the question again as to whether the predicted timetable for the demise of silicon has again been declared too early!! Ferroelectrics – a class of materials used so effectively in conjunction with silicon – certainly deserve to be here.

The chapters in Part E (Novel Materials and Selected Applications), break new ground in a number of admirable ways. Most of us are aware of, and frequently use, information recording devices such as CDs, videos, DVDs etc., but few are aware of the materials, or principles, involved. This book describes magnetic information storage materials, as well as phase-change optical recording, keeping us fully up-to-date with recent developments. The chapters also include applications such as solar cells, sensors, photoconductors, and carbon nanotubes, on which such a huge volume of work is presently being pursued worldwide. Both ends of the spectrum from research to applications are represented in chapters on molecular electronics and packaging materials.

A particular strength of this book is that it ranges from the fundamental science (Part A) through growth and characterisation of the materials (Part B) to



Prof. Arthur Willoughby
Materials Research Group,
University of Southampton,
UK

applications (Parts C–E). Virtually all the materials covered here have a wide range of applications, which is one of the reasons why this book is going to be so useful. As I indicated before, few of us will be successful in predicting the future direction and trends,

occupying the high-ground in this field in the coming decade, but this book teaches us the basic principles of materials, and leaves it to us to adapt these to the needs of tomorrow. I commend it to you most warmly.

Preface

Other handbooks in various disciplines such as electrical engineering, electronics, biomedical engineering, materials science, etc. are currently available and well used by numerous students, instructors and professionals. Most libraries have these handbook sets and each contains numerous (at least 50) chapters that cover a wide spectrum of topics within each well-defined discipline. The subject and the level of coverage appeal to both undergraduate and postgraduate students and researchers as well as to practicing professionals. The advanced topics follow introductory topics and provide ample information that is useful to all, beginners and researchers, in the field. Every few years, a new edition is brought out to update the coverage and include new topics.

There has been no similar handbook in electronic and photonic materials, and the present Springer Handbook of Electronic and Photonic Materials (SHEPM) idea grew out of a need for a handbook that covers a wide spectrum of topics in materials that today's engineers, material scientists, physicists, and chemists need. Electronic and photonic materials is a truly interdisciplinary subject that encompasses a number of traditional disciplines such as materials science, electrical engineering, chemical engineering, mechanical engineering, physics and chemistry. It is not unusual to find a mechanical engineering faculty carrying out research on electronic packaging and electrical engineers carrying out characterization measurements on semiconductors. There are only a few established university departments in electronic or photonic materials. In general, electronic materials as a "discipline" appears as a research group or as an interdisciplinary activity within a "college". One could argue that, because of the very fact that it is such an interdisciplinary field, there is a greater need to have a handbook that covers not only fundamental topics but also advanced topics; hence the present handbook.

This handbook is a comprehensive treatise on electronic and photonic materials with each chapter written by experts in the field. The handbook is aimed at senior undergraduate and graduate students, researchers and professionals working in the area of electronic, optoelectronic and photonic materials. The chapters provide the necessary background and up-to-date knowledge

in a wide range of topics. Each chapter has an introduction to the topic, many clear illustrations and numerous references. Clear explanations and illustrations make the handbook useful to all levels of researchers. All chapters are as self-contained as possible. There are both fundamental and advanced chapters to appeal to readers with different backgrounds. This is particularly important for this handbook since the subject matter is highly interdisciplinary. For example, there will be readers with a background (first degree) in chemical engineering and working on semiconductor processing who need to learn the fundamentals of semiconductors physics. Someone with a first degree in physics would need to quickly update himself on materials science concepts such as liquid phase epitaxy and so on. Difficult mathematics has been avoided and, whenever possible, the explanations have been given semiquantitatively. There is a "*Glossary of Defining Terms*" at the end of the handbook, which can serve to quickly find the definition of a term – a very necessary feature in an interdisciplinary handbook.

The editors are very grateful to all the authors for their excellent contributions and for their cooperation in delivering their manuscripts and in the various stages of production of this handbook. Sincere thanks go to Greg Franklin at Springer Boston for all his support and help throughout the long period of commissioning, acquiring the contributions and the production of the handbook. Dr. Werner Skolaut at Springer Heidelberg has very skillfully handled the myriad production issues involved in copy-editing, figure redrawing and proof preparation and correction and our sincere thanks go to him also for all his hard



Dr. Peter Capper
Materials Team Leader,
SELEX Sensors and Airborne
Systems,
Southampton, UK



Prof. Safa Kasap
Professor and Canada
Research Chair,
Electrical Engineering
Department,
University of Saskatchewan,
Canada

work in making the handbook attractive to read. He is the most dedicated and efficient editor we have come across.

It is a pleasure to thank Professor Arthur Willoughby for his many helpful suggestions that made this a better handbook. His wealth of experience as editor of the *Journal of Materials Science: Materials in Electronics* played an important role not only in selecting chapters but also in finding the right authors.

Finally, the editors wish to thank all the members of our families (Marian, Samuel and Thomas; and Nicolette) for their support and particularly their endurance during the entire project.

Peter Capper and Safa Kasap
Editors

List of Authors

Martin Abkowitz

1198 Gatestone Circle
Webster, NY 14580, USA
e-mail: mabkowitz@mailaps.org,
abkowitz@chem.chem.rochester.edu

Sadao Adachi

Gunma University
Department of Electronic Engineering,
Faculty of Engineering
Kiryu-shi 376-8515
Gunma, Japan
e-mail: adachi@el.gunma-u.ac.jp

Alfred Adams

University of Surrey
Advanced Technology Institute
Guildford, Surrey, GU2 7XH,
Surrey, UK
e-mail: alf.adams@surrey.ac.uk

Guy J. Adriaenssens

University of Leuven
Laboratorium voor Halfgeleiderfysica
Celestijnenlaan 200D
B-3001 Leuven, Belgium
e-mail: guy.adri@fys.kuleuven.ac.be

Wilfried von Ammon

Siltronic AG
Research and Development
Johannes Hess Strasse 24
84489 Burghausen, Germany
e-mail: wilfried.ammon@siltronic.com

Peter Ashburn

University of Southampton
School of Electronics and Computer Science
Southampton, S017 1BJ, UK
e-mail: pa@ecs.soton.ac.uk

Mark Auslender

Ben-Gurion University of the Negev Beer Sheva
Department of Electrical
and Computer Engineering
P.O.Box 653
Beer Sheva 84105, Israel
e-mail: marka@ee.bgu.ac.il

Darren M. Bagnall

University of Southampton
School of Electronics and Computer Science
Southampton, S017 1BJ, UK
e-mail: dmb@ecs.soton.ac.uk

Ian M. Baker

SELEX Sensors and Airborne Systems Infrared Ltd.
Southampton, Hampshire S015 OEG, UK
e-mail: ian.m.baker@selex-sas.com

Sergei Baranovskii

Philipps University Marburg
Department of Physics
Renthof 5
35032 Marburg, Germany
e-mail: baranovs@staff.uni-marburg.de

Mark Baxendale

Queen Mary, University of London
Department of Physics
Mile End Road
London, E1 4NS, UK
e-mail: m.baxendale@qmul.ac.uk

Mohammed L. Benkhedir

University of Leuven
Laboratorium voor Halfgeleiderfysica
Celestijnenlaan 200D
B-3001 Leuven, Belgium
e-mail: MohammedLoufti.Benkhedir@fys.kuleuven.ac.be

Monica Brinza

University of Leuven
Laboratorium voor Halfgeleiderfysica
Celestijnenlaan 200D
B-3001 Leuven, Belgium
e-mail: monica.brinza@fys.kuleuven.ac.be

Paul D. Brown

University of Nottingham
School of Mechanical, Materials and
Manufacturing Engineering
University Park
Nottingham, NG7 2RD, UK
e-mail: paul.brown@nottingham.ac.uk

Mike Brozel

University of Glasgow
Department of Physics and Astronomy
Kelvin Building
Glasgow, G12 8QQ, UK
e-mail: mikebrozel@beeb.net

Lukasz Brzozowski

University of Toronto
Sunnybrook and Women's Research Institute,
Imaging Research/
Department of Medical Biophysics
Research Building, 2075 Bayview Avenue
Toronto, ON, M4N 3M5, Canada
e-mail: lukbroz@sten.sunnybrook.utoronto.ca

Peter Capper

SELEX Sensors and Airborne Systems Infrared Ltd.
Materials Team Leader
Millbrook Industrial Estate, PO Box 217
Southampton, Hampshire SO15 0EG, UK
e-mail: pete.capper@selex-sas.com

Larry Comstock

San Jose State University
6574 Crystal Springs Drive
San Jose, CA 95120, USA
e-mail: Comstock@email.sjsu.edu

Ray DeCorby

University of Alberta
Department of Electrical
and Computer Engineering
7th Floor, 9107-116 Street N.W.
Edmonton, Alberta T6G 2V4, Canada
e-mail: rdecorby@trlabs.ca

M. Jamal Deen

McMaster University
Department of Electrical
and Computer Engineering (CRL 226)
1280 Main Street West
Hamilton, ON L8S 4K1, Canada
e-mail: jamal@mcmaster.ca

Leonard Dissado

The University of Leicester
Department of Engineering
University Road
Leicester, LE1 7RH, UK
e-mail: lad4@le.ac.uk

David Dunmur

University of Southampton
School of Chemistry
Southampton, SO17 1BJ, UK
e-mail: d.a.dunmur@soton.ac.uk

Lester F. Eastman

Cornell University
Department of Electrical
and Computer Engineering
425 Phillips Hall
Ithaca, NY 14853, USA
e-mail: lfe2@cornell.edu

Andy Edgar

Victoria University
School of Chemical and Physical Sciences SCPS
Kelburn Parade/PO Box 600
Wellington, New Zealand
e-mail: Andy.Edgar@vuw.ac.nz

Brian E. Foutz

Cadence Design Systems
1701 North Street, Bldg 257-3
Endicott, NY 13760, USA
e-mail: foutz@cadence.com

Mark Fox

University of Sheffield
Department of Physics and Astronomy
Hicks Building, Hounsefield Road
Sheffield, S3 7RH, UK
e-mail: mark.fox@shef.ac.uk

Darrel Frear

RF and Power Packaging Technology Development,
Freescale Semiconductor
2100 East Elliot Road
Tempe, AZ 85284, USA
e-mail: darrel.frear@freescale.com

Milan Friesel

Chalmers University of Technology
Department of Physics
Fysikgränd 3
41296 Göteborg, Sweden
e-mail: friesel@chalmers.se

Jacek Gieraltowski

Université de Bretagne Occidentale
6 Avenue Le Gorgeu, BP: 809
29285 Brest Cedex, France
e-mail: Jacek.Gieraltowski@univ-brest.fr

Yinyan Gong

Columbia University
Department of Applied Physics
and Applied Mathematics
500 W. 120th St.
New York, NY 10027, USA
e-mail: yg2002@columbia.edu

Robert D. Gould[†]

Keele University
Thin Films Laboratory, Department of Physics,
School of Chemistry and Physics
Keele, Staffordshire ST5 5BG, UK

Shlomo Hava

Ben-Gurion University of the Negev Beer Sheva
Department of Electrical
and Computer Engineering
P.O. Box 653
Beer Sheva 84105, Israel
e-mail: hava@ee.bgu.ac.il

Colin Humphreys

University of Cambridge
Department of Materials Science and Metallurgy
Pembroke Street
Cambridge, CB2 3!Z, UK
e-mail: colin.Humphreys@msm.cam.ac.uk

Stuart Irvine

University of Wales, Bangor
Department of Chemistry
Gwynedd, LL57 2UW, UK
e-mail: sjc.irvine@bangor.ac.uk

Minoru Isshiki

Tohoku University
Institute of Multidisciplinary Research
for Advanced Materials
1-1, Katahira, 2 chome, Aobaku
Sendai, 980-8577, Japan
e-mail: issshiki@tagen.tohoku.ac.jp

Robert Johanson

University of Saskatchewan
Department of Electrical Engineering
57 Campus Drive
Saskatoon, SK S7N 5A9, Canada
e-mail: johanson@engr.usask.ca

Tim Joyce

University of Liverpool
Functional Materials Research Centre,
Department of Engineering
Brownlow Hill
Liverpool, L69 3BX, UK
e-mail: t.joyce@liv.ac.uk

M. Zahangir Kabir

Concordia University
Department of Electrical and Computer Engineering
Montreal, Quebec S7N5A9, Canada
e-mail: kabir@encs.concordia.ca

Safa Kasap

University of Saskatchewan
Department of Electrical Engineering
57 Campus Drive
Saskatoon, SK S7N 5A9, Canada
e-mail: safa.kasap@usask.ca

Alexander Kolobov

National Institute of Advanced Industrial Science and Technology
Center for Applied Near-Field Optics Research
1-1-1 Higashi, Tsukuba
Ibaraki, 305-8562, Japan
e-mail: a.kolobov@aist.go.jp

Cyril Koughia

University of Saskatchewan
Department of Electrical Engineering
57 Campus Drive
Saskatoon, SK S7N 5A9, Canada
e-mail: kik486@mail.usask.ca

Igor L. Kuskovsky

Queens College, City University of New York (CUNY)
Department of Physics
65-30 Kissena Blvd.
Flushing, NY 11367, USA
e-mail: igor_kuskovsky@qc.edu

Geoffrey Luckhurst

University of Southampton
School of Chemistry
Southampton, SO17 1BJ, UK
e-mail: g.r.luckhurst@soton.ac.uk

Akihisa Matsuda

Tokyo University of Science
Research Institute for Science and Technology
2641 Yamazaki, Noda-shi
Chiba, 278-8510, Japan
e-mail: amatsuda@rs.noda.tus.ac.jp,
a.matsuda@aist.go.jp

Naomi Matsuura

Sunnybrook Health Sciences Centre
Department of Medical Biophysics,
Imaging Research
2075 Bayview Avenue
Toronto, ON M4N 3M5, Canada
e-mail: matsuura@sri.utoronto.ca

Kazuo Morigaki

University of Tokyo
C-305, Wakabadai 2-12, Inagi
Tokyo, 206-0824, Japan
e-mail: k.morigaki@yacht.ocn.ne.jp

Hadis Morkoç

Virginia Commonwealth University
Department of Electrical
and Computer Engineering
601 W. Main St., Box 843072
Richmond, VA 23284-3068, USA
e-mail: hmorkoc@vcu.edu

Winfried Mönch

Universität Duisburg-Essen
Lotharstraße 1
47048 Duisburg, Germany
e-mail: w.moench@uni-duisburg.de

Arokia Nathan

University of Waterloo
Department of Electrical
and Computer Engineering
200 University Avenue W.
Waterloo, Ontario N2L 3G1, Canada
e-mail: anathan@uwaterloo.ca

Gertrude F. Neumark

Columbia University
 Department of Applied Physics
 and Applied Mathematics
 500W 120th St., MC 4701
 New York, NY 10027, USA
 e-mail: *gfn1@columbia.edu*

Stephen K. O'Leary

University of Regina
 Faculty of Engineering
 3737 Wascana Parkway
 Regina, SK S4S 0A2, Canada
 e-mail: *stephen.oleary@uregina.ca*

Chisato Ogihara

Yamaguchi University
 Department of Applied Science
 2-16-1 Tokiwadai
 Ube, 755-8611, Japan
 e-mail: *ogihara@yamaguchi-u.ac.jp*

Fabien Pascal

Université Montpellier 2/CEM2-cc084
 Centre d'Electronique
 et de Microoptoélectronique de Montpellier
 Place E. Bataillon
 34095 Montpellier, France
 e-mail: *pascal@cem2.univ-montp2.fr*

Michael Petty

University of Durham
 Department School of Engineering
 South Road
 Durham, DH1 3LE, UK
 e-mail: *m.c.petty@durham.ac.uk*

Asim Kumar Ray

Queen Mary, University of London
 Department of Materials
 Mile End Road
 London, E1 4NS, UK
 e-mail: *a.k.ray@qmul.ac.uk*

John Rowlands

University of Toronto
 Department of Medical Biophysics
 Sunnybrook and Women's College
 Health Sciences Centre
 S656-2075 Bayview Avenue
 Toronto, ON M4N 3M5, Canada
 e-mail: *john.rowlands@sri.utoronto.ca*

Oleg Rubel

Philipps University Marburg
 Department of Physics
 and Material Sciences Center
 Renthof 5
 35032 Marburg, Germany
 e-mail: *oleg.rubel@physik.uni-marburg.de*

Harry Ruda

University of Toronto
 Materials Science and Engineering,
 Electrical and Computer Engineering
 170 College Street
 Toronto, M5S 3E4, Canada
 e-mail: *ruda@ecf.utoronto.ca*

Edward Sargent

University of Toronto
 Department of Electrical
 and Computer Engineering
 ECE, 10 King's College Road
 Toronto, M5S 3G4, Canada
 e-mail: *ted.sargent@utoronto.ca*

Peyman Servati

Ignis Innovation Inc.
 55 Culpepper Dr.
 Waterloo, Ontario N2L 5K8, Canada
 e-mail: *pservati@uwaterloo.ca*

Derek Shaw

Hull University
 Hull, HU6 7RX, UK
 e-mail: *DerekShaw1@compuserve.com*

Fumio Shimura

Shizuoka Institute of Science and Technology
Department of Materials and Life Science
2200-2 Toyosawa
Fukuroi, Shizuoka 437-8555, Japan
e-mail: shimura@ms.sist.ac.jp

Michael Shur

Rensselaer Polytechnic Institute
Department of Electrical, Computer,
and Systems Engineering
CII 9017, RPI, 110 8th Street
Troy, NY 12180, USA
e-mail: shurm@rpi.edu

Jai Singh

Charles Darwin University
School of Engineering and Logistics,
Faculty of Technology, B-41
Ellengowan Drive
Darwin, NT 0909, Australia
e-mail: jai.singh@cdu.edu.au

Tim Smeeton

Sharp Laboratories of Europe
Edmund Halley Road, Oxford Science Park
Oxford, OX4 4GB, UK
e-mail: tim.smeeton@sharp.co.uk

Boris Straumal

Russian Academy of Sciences
Institute of Solid State Physics
Institutskii prospect 15
Chernogolovka, 142432, Russia
e-mail: straumal@issp.ac.ru

Stephen Sweeney

University of Surrey
Advanced Technology Institute
Guildford, Surrey GU2 7XH, UK
e-mail: s.sweeney@surrey.ac.uk

David Sykes

Loughborough Surface Analysis Ltd.
PO Box 5016, Unit FC, Holywell Park, Ashby Road
Loughborough, LE11 3WS, UK
e-mail: d.e.sykes@lsaltd.co.uk

Keiji Tanaka

Hokkaido University
Department of Applied Physics,
Graduate School of Engineering
Kita-ku, N13 W8
Sapporo, 060-8628, Japan
e-mail: keiji@eng.hokudai.ac.jp

Charbel Tannous

Université de Bretagne Occidentale
LMB, CNRS FRE 2697
6 Avenue Le Gorgeu, BP: 809
29285 Brest Cedex, France
e-mail: tannous@univ-brest.fr

Ali Teke

Balikesir University
Department of Physics, Faculty of Art and Science
Balikesir, 10100, Turkey
e-mail: ateke@balikesir.edu.tr

Junji Tominaga

National Institute of Advanced Industrial
Science and Technology, AIST
Center for Applied Near-Field Optics Research,
CAN-FOR
Tsukuba Central 4 1-1-1 Higashi
Tsukuba, 3-5-8562, Japan
e-mail: j-tomonaga@aist.go.jp

Dan Tonchev

University of Saskatchewan
Department of Electrical Engineering
57 Campus Drive
Saskatoon, SK S7N 5A9, Canada
e-mail: dan.tonchev@usask.ca

Harry L. Tuller

Massachusetts Institute of Technology
Department of Materials Science and Engineering,
Crystal Physics and Electroceramics Laboratory
77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: tuller@mit.edu

Qamar-ul Wahab

Linköping University
Department of Physics,
Chemistry, and Biology (IFM)
SE-581 83 Linköping, Sweden
e-mail: quw@ifm.liu.se

Robert M. Wallace

University of Texas at Dallas
Department of Electrical Engineering
M.S. EC 33, P.O.Box 830688
Richardson, TX 75083, USA
e-mail: rmwallace@utdallas.edu

Jifeng Wang

Tohoku University
Institute of Multidisciplinary Research
for Advanced Materials
1-1, Katahira, 2 Chome, Aobaku
Sendai, 980-8577, Japan
e-mail: wang@tagen.tohoku.ac.jp

David S. Weiss

NexPress Solutions, Inc.
2600 Manitou Road
Rochester, NY 14653-4180, USA
e-mail: David_Weiss@Nexpress.com

Rainer Wesche

Swiss Federal Institute of Technology
Centre de Recherches en Physique des Plasmas
CRPP (c/o Paul Scherrer Institute), WMHA/C31,
Villigen PS
Lausanne, CH-5232, Switzerland
e-mail: rainer.wesche@psi.ch

Roger Whatmore

Tyndall National Institute
Lee Maltings, Cork, Ireland
e-mail: roger.whatmore@tyndall.ie

Neil White

University of Southampton
School of Electronics and Computer Science
Mountbatten Building
Highfield, Southampton SO17 1BJ, UK
e-mail: nmw@ecs.soton.ac.uk

Magnus Willander

University of Gothenburg
Department of Physics
SE-412 96 Göteborg, Sweden
e-mail: mwi@fy.chalmers.se

Jan Willekens

University of Leuven
Laboratorium voor Halfgeleiderfysica
Celestijnenlaan 200D
B-3001 Leuven, Belgium
e-mail: jan.willekens@kc.kuleuven.ac.be

Contents

List of Abbreviations	XXIX
------------------------------------	-------------

Introduction

1 Perspectives on Electronic and Optoelectronic Materials.....	3
1.1 The Early Years	4
1.2 The Silicon Age	4
1.3 The Compound Semiconductors.....	8
1.4 From Faraday to Today	14
References.....	14

Part A Fundamental Properties

2 Electrical Conduction in Metals and Semiconductors.....	19
2.1 Fundamentals: Drift Velocity, Mobility and Conductivity	20
2.2 Matthiessen's Rule	22
2.3 Resistivity of Metals.....	23
2.4 Solid Solutions and Nordheim's Rule	26
2.5 Carrier Scattering in Semiconductors	28
2.6 The Boltzmann Transport Equation.....	29
2.7 Resistivity of Thin Polycrystalline Films	30
2.8 Inhomogeneous Media. Effective Media Approximation	32
2.9 The Hall Effect	35
2.10 High Electric Field Transport.....	37
2.11 Avalanche	38
2.12 Two-Dimensional Electron Gas.....	39
2.13 One Dimensional Conductance	41
2.14 The Quantum Hall Effect	42
References.....	44
3 Optical Properties of Electronic Materials: Fundamentals and Characterization	47
3.1 Optical Constants.....	47
3.2 Refractive Index	50
3.3 Optical Absorption.....	53
3.4 Thin Film Optics.....	70
3.5 Optical Materials	74
References.....	76

4	Magnetic Properties of Electronic Materials	79
4.1	Traditional Magnetism	81
4.2	Unconventional Magnetism	93
	References	99
5	Defects in Monocrystalline Silicon	101
5.1	Technological Impact of Intrinsic Point Defects Aggregates	102
5.2	Thermophysical Properties of Intrinsic Point Defects	103
5.3	Aggregates of Intrinsic Point Defects	104
5.4	Formation of OSF Ring	115
	References	117
6	Diffusion in Semiconductors	121
6.1	Basic Concepts	122
6.2	Diffusion Mechanisms	122
6.3	Diffusion Regimes	123
6.4	Internal Electric Fields	126
6.5	Measurement of Diffusion Coefficients	126
6.6	Hydrogen in Semiconductors	127
6.7	Diffusion in Group IV Semiconductors	128
6.8	Diffusion in III–V Compounds	130
6.9	Diffusion in II–VI Compounds	131
6.10	Conclusions	133
6.11	General Reading and References	133
	References	133
7	Photoconductivity in Materials Research	137
7.1	Steady State Photoconductivity Methods	138
7.2	Transient Photoconductivity Experiments	142
	References	146
8	Electronic Properties of Semiconductor Interfaces	147
8.1	Experimental Database	149
8.2	IFIGS-and-Electronegativity Theory	153
8.3	Comparison of Experiment and Theory	155
8.4	Final Remarks	159
	References	159
9	Charge Transport in Disordered Materials	161
9.1	General Remarks on Charge Transport in Disordered Materials	163
9.2	Charge Transport in Disordered Materials via Extended States	167
9.3	Hopping Charge Transport in Disordered Materials via Localized States	169
9.4	Concluding Remarks	184
	References	185

10 Dielectric Response	187
10.1 Definition of Dielectric Response	188
10.2 Frequency-Dependent Linear Responses	190
10.3 Information Contained in the Relaxation Response	196
10.4 Charge Transport	208
10.5 A Few Final Comments	211
References	211
11 Ionic Conduction and Applications	213
11.1 Conduction in Ionic Solids	214
11.2 Fast Ion Conduction	216
11.3 Mixed Ionic–Electronic Conduction	221
11.4 Applications	223
11.5 Future Trends	226
References	226
 Part B Growth and Characterization	
12 Bulk Crystal Growth – Methods and Materials	231
12.1 History	232
12.2 Techniques	233
12.3 Materials Grown	240
12.4 Conclusions	251
References	251
13 Single-Crystal Silicon: Growth and Properties	255
13.1 Overview	256
13.2 Starting Materials	257
13.3 Single-Crystal Growth	258
13.4 New Crystal Growth Methods	266
References	268
14 Epitaxial Crystal Growth: Methods and Materials	271
14.1 Liquid-Phase Epitaxy (LPE)	271
14.2 Metalorganic Chemical Vapor Deposition (MOCVD)	280
14.3 Molecular Beam Epitaxy (MBE)	290
References	299
15 Narrow-Bandgap II–VI Semiconductors: Growth	303
15.1 Bulk Growth Techniques	304
15.2 Liquid-Phase Epitaxy (LPE)	308
15.3 Metalorganic Vapor Phase Epitaxy (MOVPE)	312
15.4 Molecular Beam Epitaxy (MBE)	317
15.5 Alternatives to CMT	320
References	321

16 Wide-Bandgap II–VI Semiconductors: Growth and Properties	325
16.1 Crystal Properties	326
16.2 Epitaxial Growth	328
16.3 Bulk Crystal Growth	333
16.4 Conclusions	339
References	340
17 Structural Characterization	343
17.1 Radiation–Material Interactions	344
17.2 Particle–Material Interactions	345
17.3 X-Ray Diffraction	348
17.4 Optics, Imaging and Electron Diffraction	351
17.5 Characterizing Functional Activity	362
17.6 Sample Preparation	362
17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials	364
17.8 Concluding Remarks	370
References	370
18 Surface Chemical Analysis	373
18.1 Electron Spectroscopy	373
18.2 Glow–Discharge Spectroscopies (GDOES and GDMS)	376
18.3 Secondary Ion Mass Spectrometry (SIMS)	377
18.4 Conclusion	384
19 Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications	385
19.1 Heat Capacity	386
19.2 Thermal Conductivity	391
19.3 Thermal Expansion	396
19.4 Enthalpic Thermal Properties	398
19.5 Temperature–Modulated DSC (TMDSC)	403
References	406
20 Electrical Characterization of Semiconductor Materials and Devices	409
20.1 Resistivity	410
20.2 Hall Effect	418
20.3 Capacitance–Voltage Measurements	421
20.4 Current–Voltage Measurements	426
20.5 Charge Pumping	428
20.6 Low–Frequency Noise	430
20.7 Deep–Level Transient Spectroscopy	434
References	436

Part C Materials for Electronics

21 Single-Crystal Silicon: Electrical and Optical Properties	441
21.1 Silicon Basics	441
21.2 Electrical Properties	451
21.3 Optical Properties	472
References.....	478
22 Silicon–Germanium: Properties, Growth and Applications	481
22.1 Physical Properties of Silicon–Germanium	482
22.2 Optical Properties of SiGe	488
22.3 Growth of Silicon–Germanium	492
22.4 Polycrystalline Silicon–Germanium	494
References.....	497
23 Gallium Arsenide	499
23.1 Bulk Growth of GaAs	502
23.2 Epitaxial Growth of GaAs	507
23.3 Diffusion in Gallium Arsenide	511
23.4 Ion Implantation into GaAs	513
23.5 Crystalline Defects in GaAs	514
23.6 Impurity and Defect Analysis of GaAs (Chemical)	517
23.7 Impurity and Defect Analysis of GaAs (Electrical)	518
23.8 Impurity and Defect Analysis of GaAs (Optical)	521
23.9 Assessment of Complex Heterostructures	522
23.10 Electrical Contacts to GaAs	524
23.11 Devices Based on GaAs (Microwave).....	524
23.12 Devices based on GaAs (Electro-optical)	527
23.13 Other Uses for GaAs	532
23.14 Conclusions.....	532
References.....	533
24 High-Temperature Electronic Materials: Silicon Carbide and Diamond	537
24.1 Material Properties and Preparation	540
24.2 Electronic Devices	547
24.3 Summary	557
References.....	558
25 Amorphous Semiconductors: Structure, Optical, and Electrical Properties.....	565
25.1 Electronic States	565
25.2 Structural Properties	568
25.3 Optical Properties	570
25.4 Electrical Properties	573
25.5 Light-Induced Phenomena	575
25.6 Nanosized Amorphous Structure.....	577
References.....	578

26 Amorphous and Microcrystalline Silicon	581
26.1 Reactions in SiH_4 and SiH_4/H_2 Plasmas	581
26.2 Film Growth on a Surface	583
26.3 Defect Density Determination for a-Si:H and $\mu\text{c-Si:H}$	589
26.4 Device Applications	590
26.5 Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells.....	591
26.6 Summary	594
References.....	594
27 Ferroelectric Materials	597
27.1 Ferroelectric Materials	601
27.2 Ferroelectric Materials Fabrication Technology	608
27.3 Ferroelectric Applications.....	616
References.....	622
28 Dielectric Materials for Microelectronics	625
28.1 Gate Dielectrics	630
28.2 Isolation Dielectrics	647
28.3 Capacitor Dielectrics	647
28.4 Interconnect Dielectrics	651
28.5 Summary	653
References.....	653
29 Thin Films	659
29.1 Deposition Methods	661
29.2 Structure	682
29.3 Properties	692
29.4 Concluding Remarks	708
References.....	711
30 Thick Films	717
30.1 Thick Film Processing.....	718
30.2 Substrates.....	720
30.3 Thick Film Materials.....	721
30.4 Components and Assembly	724
30.5 Sensors	728
References.....	731

Part D Materials for Optoelectronics and Photonics

31 III-V Ternary and Quaternary Compounds	735
31.1 Introduction to III-V Ternary and Quaternary Compounds	735
31.2 Interpolation Scheme	736
31.3 Structural Parameters	737
31.4 Mechanical, Elastic and Lattice Vibronic Properties.....	739

31.5	Thermal Properties	741
31.6	Energy Band Parameters	743
31.7	Optical Properties	748
31.8	Carrier Transport Properties	750
	References	751
32	Group III Nitrides	753
32.1	Crystal Structures of Nitrides	755
32.2	Lattice Parameters of Nitrides	756
32.3	Mechanical Properties of Nitrides	757
32.4	Thermal Properties of Nitrides	761
32.5	Electrical Properties of Nitrides	766
32.6	Optical Properties of Nitrides	777
32.7	Properties of Nitride Alloys	791
32.8	Summary and Conclusions	794
	References	795
33	Electron Transport Within the III–V Nitride Semiconductors, GaN, AlN, and InN: A Monte Carlo Analysis	805
33.1	Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach	806
33.2	Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN	810
33.3	Electron Transport Within III–V Nitride Semiconductors: A Review	822
33.4	Conclusions	826
	References	826
34	II–IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe	829
34.1	Background	829
34.2	Solar Cells	829
34.3	Radiation Detectors	834
34.4	Conclusions	840
	References	840
35	Doping Aspects of Zn-Based Wide-Band-Gap Semiconductors	843
35.1	ZnSe	843
35.2	ZnBeSe	848
35.3	ZnO	849
	References	851
36	II–VI Narrow-Bandgap Semiconductors for Optoelectronics	855
36.1	Applications and Sensor Design	858
36.2	Photoconductive Detectors in HgCdTe and Related Alloys	860
36.3	SPRITE Detectors	864
36.4	Photoconductive Detectors in Closely Related Alloys	866
36.5	Conclusions on Photoconductive HgCdTe Detectors	867
36.6	Photovoltaic Devices in HgCdTe	867

36.7	Emission Devices in II–VI Semiconductors	882
36.8	Potential for Reduced-Dimensionality HgTe–CdTe	883
	References.....	883
37	Optoelectronic Devices and Materials	887
37.1	Introduction to Optoelectronic Devices	888
37.2	Light-Emitting Diodes and Semiconductor Lasers	890
37.3	Single-Mode Lasers	904
37.4	Optical Amplifiers	906
37.5	Modulators	907
37.6	Photodetectors.....	911
37.7	Conclusions.....	914
	References.....	915
38	Liquid Crystals	917
38.1	Introduction to Liquid Crystals	917
38.2	The Basic Physics of Liquid Crystals	924
38.3	Liquid-Crystal Devices	931
38.4	Materials for Displays	940
	References.....	949
39	Organic Photoconductors	953
39.1	Chester Carlson and Xerography	954
39.2	Operational Considerations and Critical Materials Properties	956
39.3	OPC Characterization	965
39.4	OPC Architecture and Composition	967
39.5	Photoreceptor Fabrication	976
39.6	Summary	977
	References.....	978
40	Luminescent Materials	983
40.1	Luminescent Centres.....	985
40.2	Interaction with the Lattice	987
40.3	Thermally Stimulated Luminescence.....	989
40.4	Optically (Photo-)Stimulated Luminescence	990
40.5	Experimental Techniques – Photoluminescence.....	991
40.6	Applications.....	992
40.7	Representative Phosphors	995
	References.....	995
41	Nano-Engineered Tunable Photonic Crystals in the Near-IR and Visible Electromagnetic Spectrum	997
41.1	PC Overview	998
41.2	Traditional Fabrication Methodologies for Static PCs	1001
41.3	Tunable PCs.....	1011
41.4	Summary and Conclusions	1014
	References.....	1015

42 Quantum Wells, Superlattices, and Band-Gap Engineering	1021
42.1 Principles of Band-Gap Engineering and Quantum Confinement	1022
42.2 Optoelectronic Properties of Quantum-Confined Structures	1024
42.3 Emitters	1032
42.4 Detectors	1034
42.5 Modulators	1036
42.6 Future Directions	1037
42.7 Conclusions	1038
References	1038
43 Glasses for Photonic Integration	1041
43.1 Main Attributes of Glasses as Photonic Materials	1042
43.2 Glasses for Integrated Optics	1050
43.3 Laser Glasses for Integrated Light Sources	1053
43.4 Summary	1057
References	1059
44 Optical Nonlinearity in Photonic Glasses	1063
44.1 Third-Order Nonlinearity in Homogeneous Glass	1064
44.2 Second-Order Nonlinearity in Poled Glass	1069
44.3 Particle-Embedded Systems	1070
44.4 Photoinduced Phenomena	1071
44.5 Summary	1072
References	1072
45 Nonlinear Optoelectronic Materials	1075
45.1 Background	1075
45.2 Illumination-Dependent Refractive Index and Nonlinear Figures of Merit (FOM)	1077
45.3 Bulk and Multi-Quantum-Well (MQW) Inorganic Crystalline Semiconductors	1080
45.4 Organic Materials	1084
45.5 Nanocrystals	1087
45.6 Other Nonlinear Materials	1088
45.7 Conclusions	1089
References	1089

Part E Novel Materials and Selected Applications

46 Solar Cells and Photovoltaics	1095
46.1 Figures of Merit for Solar Cells	1096
46.2 Crystalline Silicon	1098
46.3 Amorphous Silicon	1100
46.4 GaAs Solar Cells	1101
46.5 CdTe Thin-Film Solar Cells	1102

46.6	CuInGaSe ₂ (CIGS) Thin-Film Solar Cells	1103
46.7	Conclusions	1104
	References	1105
47	Silicon on Mechanically Flexible Substrates for Large-Area Electronics	1107
47.1	a-Si:H TFTs on Flexible Substrates	1108
47.2	Field-Effect Transport in Amorphous Films	1108
47.3	Electronic Transport Under Mechanical Stress	1113
	References	1118
48	Photoconductors for X-Ray Image Detectors	1121
48.1	X-Ray Photoconductors	1123
48.2	Metrics of Detector Performance	1131
48.3	Conclusion	1136
	References	1136
49	Phase-Change Optical Recording	1139
49.1	Digital Versatile Discs (DVDs)	1140
49.2	Super-RENS Discs	1144
49.3	In Lieu of Conclusion	1145
	References	1145
50	Carbon Nanotubes and Bucky Materials	1147
50.1	Carbon Nanotubes	1147
50.2	Bucky Materials	1153
	References	1153
51	Magnetic Information-Storage Materials	1155
51.1	Magnetic Recording Technology	1156
51.2	Magnetic Random-Access Memory	1185
51.3	Extraordinary Magnetoresistance (EMR)	1189
51.4	Summary	1189
	References	1189
52	High-Temperature Superconductors	1193
52.1	The Superconducting State	1195
52.2	Cuprate High- T_c Superconductors: An Overview	1202
52.3	Physical Properties of Cuprate Superconductors	1207
52.4	Superconducting Films	1212
52.5	The Special Case of MgB ₂	1214
52.6	Summary	1216
	References	1216
53	Molecular Electronics	1219
53.1	Electrically Conductive Organic Compounds	1220
53.2	Materials	1223
53.3	Plastic Electronics	1225

53.4	Molecular-Scale Electronics	1229
53.5	DNA Electronics	1235
53.6	Conclusions	1236
	References	1237
54	Organic Materials for Chemical Sensing	1241
54.1	Analyte Requirements	1242
54.2	Brief Review of Inorganic Materials	1243
54.3	Macrocyclic Compounds for Sensing	1245
54.4	Sensing with Phthalocyanine and Porphyrin	1250
54.5	Polymeric Materials	1255
54.6	Cavitand Molecules	1259
54.7	Concluding Remarks	1261
	References	1262
55	Packaging Materials	1267
55.1	Package Applications	1268
55.2	The Materials Challenge of Electronic Packaging	1269
55.3	Materials Coefficient of Thermal Expansion	1272
55.4	Wirebond Materials	1272
55.5	Solder Interconnects	1273
55.6	Substrates	1278
55.7	Underfill and Encapsulants	1280
55.8	Electrically Conductive Adhesives (ECAs)	1281
55.9	Thermal Issues	1283
55.10	Summary	1284
	References	1285
	Acknowledgements	1287
	About the Authors	1291
	Detailed Contents	1307
	Glossary of Defining Terms	1333
	Subject Index	1367

List of Abbreviations

2DEG two-dimensional electron gas

A

AC alternating current
 ACCUFET accumulation-mode MOSFET
 ACRT accelerated crucible rotation technique
 AEM analytical electron microscopes
 AES Auger electron spectroscopy
 AFM atomic force microscopy
 ALD atomic-layer deposition
 ALE atomic-layer epitaxy
 AMA active matrix array
 AMFPI active matrix flat-panel imaging
 AMOLED amorphous organic light-emitting diode
 APD avalanche photodiode

B

b.c.c. body-centered cubic
 BEEM ballistic-electron-emission microscopy
 BEP beam effective pressure
 BH buried-heterostructure
 BH Brooks–Herring
 BJT bipolar junction transistor
 BTEX m-xylene
 BZ Brillouin zone

C

CAIBE chemically assisted ion beam etching
 CB conduction band
 CBE chemical beam epitaxy
 CBED convergent beam electron diffraction
 CC constant current
 CCD charge-coupled device
 CCZ continuous-charging Czochralski
 CFLPE container-free liquid phase epitaxy
 CKR cross Kelvin resistor
 CL cathodoluminescence
 CMOS complementary metal-oxide-semiconductor
 CNR carrier-to-noise ratio
 COP crystal-originated particle
 CP charge pumping
 CPM constant-photocurrent method
 CR computed radiography
 CR-DLTS computed radiography deep level transient spectroscopy
 CRA cast recrystallize anneal
 CTE coefficient of thermal expansion
 CTO chromium(III) trioxalate

CuPc copper phthalocyanine
 CuTTBPc tetra-tert-butyl phthalocyanine
 CV chemical vapor
 CVD chemical vapor deposition
 CVT chemical vapor transport
 CZ Czochralski
 CZT cadmium zinc telluride

D

DA Drude approximation
 DAG direct alloy growth
 DBP dual-beam photoconductivity
 DC direct current
 DCPBH double-channel planar buried heterostructure
 DET diethyl telluride
 DFB distributed feedback
 DH double heterostructure
 DIL dual-in-line
 DIPTe diisopropyltellurium
 DLC diamond-like carbon
 DLHJ double-layer heterojunction
 DLTS deep level transient spectroscopy
 DMCd dimethyl cadmium
 DMF dimethylformamide
 DMOSFET double-diffused MOSFET
 DMS dilute magnetic semiconductors
 DMSO dimethylsulfoxide
 DMZn dimethylzinc
 DOS density of states
 DQE detective quantum efficiency
 DSIMS dynamic secondary ion mass spectrometry
 DTBSe ditertiarybutylselenide
 DUT device under test
 DVD digital versatile disk
 DWDM dense wavelength-division multiplexing
 DXD double-crystal X-ray diffraction

E

EBIC electron beam induced conductivity
 ED electrodeposition
 EDFA erbium-doped fiber amplifier
 EELS electron energy loss spectroscopy
 EFG film-fed growth
 EHP electron–hole pairs
 ELO epitaxial lateral overgrowth
 ELOG epitaxial layer overgrowth
 EM electromagnetic
 EMA effective media approximation

ENDOR electron–nuclear double resonance
 EPD etch pit density
 EPR electron paramagnetic resonance
 ESR electron spin resonance spectroscopy
 EXAFS extended X-ray absorption fine structure

F

FCA free-carrier absorption
 f.c.c. face-centered cubic
 FET field effect transistor
 FIB focused ion beam
 FM Frank–van der Merwe
 FPA focal plane arrays
 FPD flow pattern defect
 FTIR Fourier transform infrared
 FWHM full-width at half-maximum
 FZ floating zone

G

GDA generalized Drude approximation
 GDMS glow discharge mass spectrometry
 GDOES glow discharge optical emission spectroscopy
 GF gradient freeze
 GMR giant magnetoresistance
 GOI gate oxide integrity
 GRIN graded refractive index
 GSMBE gas-source molecular beam epitaxy
 GTO gate turn-off

H

HAADF high-angle annular dark field
 HB horizontal Bridgman
 HBT hetero-junction bipolar transistor
 HDC horizontal directional solidification crystallization
 HEMT high electron mobility transistor
 HF high-frequency
 HOD highly oriented diamond
 HOLZ high-order Laue zone
 HPc phthalocyanine
 HPHT high-pressure high-temperature
 HRXRD high-resolution X-ray diffraction
 HTCVD high-temperature CVD
 HVDC high-voltage DC
 HWE hot-wall epitaxy

I

IC integrated circuit
 ICTS isothermal capacitance transient spectroscopy
 IDE interdigitated electrodes

IFIGS interface-induced gap states
 IFTOF interrupted field time-of-flight
 IGBT insulated gate bipolar transistor
 IMP interdiffused multilayer process
 IPEYS internal photoemission yield spectroscopy
 IR infrared
 ITO indium-tin-oxide

J

JBS junction barrier Schottky
 JFET junction field-effect transistors
 JO Judd–Ofelt

K

KCR Kelvin contact resistance
 KKR Kramers–Kronig relation
 KLN $\text{K}_3\text{Li}_2\text{Nb}_5\text{O}_{12}$
 KTPO KTiOPO_4

L

LB Langmuir–Blodgett
 LD laser diodes
 LD lucky drift
 LDD lightly doped drain
 LEC liquid-encapsulated Czochralski
 LED light-emitting diodes
 LEIS low-energy ion scattering
 LEL lower explosive limit
 LF low-frequency
 LLS laser light scattering
 LMA law of mass action
 LO longitudinal optical
 LPE liquid phase epitaxy
 LSTD laser light scattering tomography defect
 LVM localized vibrational mode

M

MBE molecular beam epitaxy
 MCCZ magnetic field applied continuous Czochralski
 MCT mercury cadmium telluride
 MCZ magnetic field applied Czochralski
 MD molecular dynamics
 MEED medium-energy electron diffraction
 MEM micro-electromechanical systems
 MESFET metal-semiconductor field-effect transistor
 MFC mass flow controllers
 MIGS metal-induced gap states
 ML monolayer
 MLHJ multilayer heterojunction
 MOCVD metal-organic chemical vapor deposition
 MODFET modulation-doped field effect transistor

MOMBE	metalorganic molecular beam epitaxy
MOS	metal/oxide/semiconductor
MOSFET	metal/oxide/semiconductor field effect transistor
MOVPE	metalorganic vapor phase epitaxy
MPc	metallophthalocyanine
MPC	modulated photoconductivity
MPCVD	microwave plasma chemical deposition
MQW	multiple quantum well
MR	magnetoresistivity
MS	metal–semiconductor
MSRD	mean-square relative displacement
MTF	modulation transfer function
MWIR	medium-wavelength infrared

N

NDR	negative differential resistance
NEA	negative electron affinity
NeXT	nonthermal energy exploration telescope
NMOS	n-type-channel metal–oxide–semiconductor
NMP	N-methylpyrrolidone
NMR	nuclear magnetic resonance
NNH	nearest-neighbor hopping
NSA	naphthalene-1,5-disulfonic acid
NTC	negative temperature coefficient
NTD	neutron transmutation doping

O

OLED	organic light-emitting diode
OSF	oxidation-induced stacking fault
OSL	optically stimulated luminescence
OZM	overlap zone melting

P

PAE	power added efficiency
PAni	polyaniline
pBN	pyrolytic boron nitride
Pc	phthalocyanine
PC	photoconductive
PCA	principal component analysis
PCB	printed circuit board
PDMA	poly(methylmethacrylate)/poly(decyl methacrylate)
PDP	plasma display panels
PDS	photothermal deflection spectroscopy
PE	polysilicon emitter
PE BJT	polysilicon emitter bipolar junction transistor
PECVD	plasma-enhanced chemical vapor deposition
PEN	polyethylene naphthalate
PES	photoemission spectroscopy
PET	positron emission tomography
pHEMT	pseudomorphic HEMT

PL	photoluminescence
PM	particulate matter
PMMA	poly(methyl-methacrylate)
POT	poly(<i>n</i> -octyl)thiophene
ppb	parts per billion
ppm	parts per million
PPS	polyphenylsulfide
PPY	polypyrrole
PQT-12	poly[5,5'-bis(3-alkyl-2-thienyl)-2,2'-bithiophene]
PRT	platinum resistance thermometers
PSt	polystyrene
PTC	positive temperature coefficient
PTIS	photothermal ionisation spectroscopy
PTS	1,1-dioxo-2-(4-methylphenyl)-6-phenyl-4-(dicyanomethylidene)thiopyran
PTV	polythienylene vinylene
PV	photovoltaic
PVD	physical vapor transport
PVDF	polyvinylidene fluoride
PVK	polyvinylcarbazole
PVT	physical vapor transport
PZT	lead zirconate titanate

Q

QA	quench anneal
QCL	quantum cascade laser
QCSE	quantum-confined Stark effect
QD	quantum dot
QHE	quantum Hall effect
QW	quantum well

R

RAIRS	reflection adsorption infrared spectroscopy
RBS	Rutherford backscattering
RCLED	resonant-cavity light-emitting diode
RDF	radial distribution function
RDS	reflection difference spectroscopy
RE	rare earth
RENS	resolution near-field structure
RF	radio frequency
RG	recombination–generation
RH	relative humidity
RHEED	reflection high-energy electron diffraction
RIE	reactive-ion etching
RIU	refractive index units
RTA	rapid thermal annealing
RTD	resistance temperature devices
RTS	random telegraph signal

S

SA	self-assembly
SAM	self-assembled monolayers

SAW	surface acoustic wave
SAXS	small-angle X-ray scattering
SCH	separate confinement heterojunction
SCVT	seeded chemical vapor transport
SE	spontaneous emission
SEM	scanning electron microscope
SIMS	secondary ion mass spectrometry
SIPBH	semi-insulating planar buried heterostructure
SIT	static induction transistors
SK	Stranski–Krastanov
SNR	signal-to-noise ratio
SO	small outline
SOA	semiconductor optical amplifier
SOC	system-on-a-chip
SOFC	solid oxide fuel cells
SOI	silicon-on-insulator
SP	screen printing
SPECT	single-photon emission computed tomography
SPR	surface plasmon resonance
SPVT	seeded physical vapor transport
SQW	single quantum wells
SSIMS	static secondary ion mass spectrometry
SSPC	steady-state photoconductivity
SSR	solid-state recrystallisation
SSRM	scanning spreading resistance microscopy
STHM	sublimation traveling heater method
SVP	saturated vapor pressure
SWIR	short-wavelength infrared

T

TAB	tab automated bonding
TBA	tertiarybutylarsine
TBP	tertiarybutylphosphine
TCE	thermal coefficient of expansion
TCNQ	tetracyanoquinodimethane
TCR	temperature coefficient of resistance
TCRI	temperature coefficient of refractive index
TDCM	time-domain charge measurement
TE	transverse electric
TED	transient enhanced diffusion
TED	transmission electron diffraction
TEGa	triethylgallium
TEM	transmission electron microscope
TEN	triethylamine
TFT	thin-film transistors
THM	traveling heater method
TL	thermoluminescence
TLHJ	triple-layer graded heterojunction
TLM	transmission line measurement
TM	transverse magnetic

TMA	trimethyl-aluminum
TMG	trimethyl-gallium
TMI	trimethyl-indium
TMSb	trimethylantimony
TO	transverse optical
TOF	time of flight
ToFSIMS	time of flight SIMS
TPC	transient photoconductivity
TPV	thermophotovoltaic
TSC	thermally stimulated current
TSL	thermally stimulated luminescence

U

ULSI	ultra-large-scale integration
UMOSFET	U-shaped-trench MOSFET
UPS	uninterrupted power systems
UV	ultraviolet

V

VAP	valence-alternation pairs
VB	valence band
VCSEL	vertical-cavity surface-emitting laser
VCZ	vapor-pressure-controlled Czochralski
VD	vapor deposition
VFE	vector flow epitaxy
VFET	vacuum field-effect transistor
VGF	vertical gradient freeze
VIS	visible
VOC	volatile organic compounds
VPE	vapor phase epitaxy
VRH	variable-range hopping
VUVG	vertical unseeded vapor growth
VW	Volmer–Weber

W

WDX	wavelength dispersive X-ray
WXI	wide-band X-ray imager

X

XAFS	X-ray absorption fine-structure
XANES	X-ray absorption near-edge structure
XEBIT	X-ray-sensitive electron-beam image tube
XPS	X-ray photon spectroscopy
XRD	X-ray diffraction
XRSP	X-ray storage phosphor

Y

YSZ	yttrium-stabilized zirconia
-----	-----------------------------

Introduction

Introduction

1 Perspectives on Electronic and Optoelectronic Materials

Tim Smeeton, Oxford, UK

Colin Humphreys, Cambridge, UK

Perspectives

1. Perspectives on Electronic and Optoelectronic Materials

This opening chapter will concentrate on the changes in the world of semiconducting materials and devices over the latter half of the twentieth century. Within this field we have chosen to concentrate on a few developments and cannot claim to cover all of the major areas. What we plan to do is give a sense of perspective of how the science and technology of these materials has come to its current state and to present a brief overview of why certain materials are chosen for particular device applications.

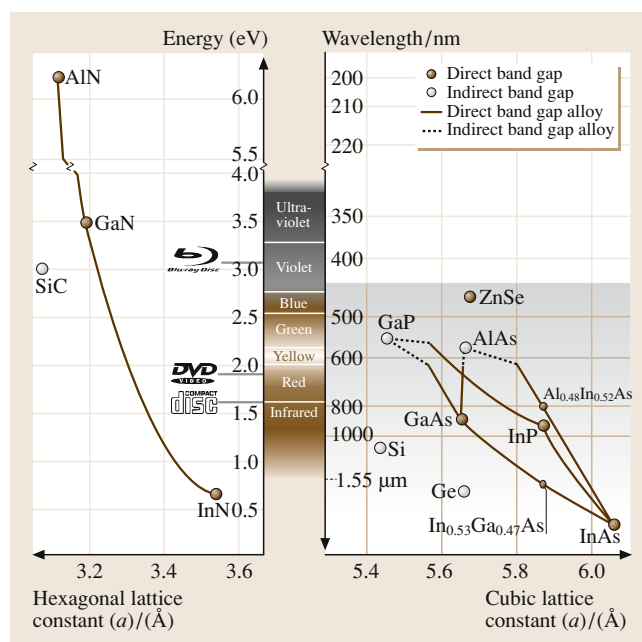
We start by identifying some of the earliest developments in our understanding of electronic materials; follow the development of silicon technology from the first demonstration of the transistor through to today's integrated circuit; track some of the key electronic and optoelectronic uses of the conventional III–V semiconductors; and end with a review of the last decade's explosion of

1.1	The Early Years	4
1.2	The Silicon Age	4
1.2.1	The Transistor and Early Semiconductor Materials Development	4
1.2.2	The Integrated Circuit	6
1.3	The Compound Semiconductors	8
1.3.1	High Speed Electronics	9
1.3.2	Light Emitting Devices	10
1.3.3	The III–Nitrides	12
1.4	From Faraday to Today	14
	References	14

interest in the III–nitride materials. The band gaps of the semiconductors encountered in this chapter are shown in Fig. 1.1 – a figure which will be frequently referred to in explaining the choice of materials for specific applications.

We sometimes forget how remarkable electronic and optoelectronic materials are. Take the light emitting diode (LED) as an example: an electric current is passed through nothing more than a tiny stack of layers of slightly different compounds and brilliant coloured light is emitted. Of course the stack in question is carefully designed, and there are theories to explain the behaviour, but that should not detract from the initial moment of wonderment that it works at all! Devices from LEDs through to microprocessors containing tens of millions of transistors influence life at the start of this twenty first century to an incalculable extent. These complex devices are only available today because an array of electronic materials have been developed using increasingly sophisticated research methods over the past six decades.

Fig. 1.1 Parameter perspective: band gaps and lattice parameters of selected semiconductors discussed in the text. The important wavelengths for optical storage (CD, DVD and Blu-Ray) and the 1.55 μm used for efficient data transmission through optical fibres are labelled



Most semiconductor-based appliances which affect us every day are made using silicon, but many key devices depend on a number of different compound semiconductors. For example GaP-based LEDs in digital displays; GaAs-based HEMTs which operate in our satellite television receivers; AlGaAs and AlInGaP lasers in our CD and DVD players; and the InP-based lasers delivering the internet and telecommunications along optic fibres.

None of these devices could be manufactured without a basic understanding (either fundamental or

empirical) of the materials science of the components. At the same time the overwhelming reason for scientific study of the materials is to elicit some improvement in the performance of the devices based on them. The rest of this book concentrates on the *materials* more than their *devices* but to give some idea of how a world largely untouched by electronic materials in the 1940s has become so changed by them we will consider the developments in the two fields in parallel in this chapter. Often they are inseparable anyway.

1.1 The Early Years

The exploitation of electronic materials in solid-state devices principally occurred in the second half of the twentieth century but the first serendipitous observations of semiconducting behaviour took place somewhat earlier than this. In 1833, *Faraday* found that silver sulphide exhibited a negative temperature coefficient [1.1]. This property of a decrease in electrical resistivity with increasing temperature was to be deployed in thermistor components a century later. In the 1870s scientists discovered and experimented with the photoconductivity (decreased resistivity of a material under incident light) of selenium [1.2, 3]. Amorphous selenium was to be used for this very property in the first Xerox copying machines of the 1950s.

While these discoveries had limited immediate impact on scientific understanding, more critical progress was made such as Hall's 1879 discovery of what was to become known as the Hall Effect. The discovery of quantum mechanics was of fundamental importance for

our understanding of semiconductors. Based on the advances in quantum theory in the early 1900s a successful theory to explain semiconductor behaviour was formulated in 1931 [1.4, 5]. However, the semiconductors of the 1930s were too impure to allow the theory to be compared with experiment. For example it was believed at the time that silicon, which was to become the archetypal semiconducting material, was a substance belonging to a group of materials which were "good metallic conductors in the pure state and . . . therefore to be classed as metals" [1.5]!

However, a solid theoretical understanding of semiconductors was in place by the 1940s. Hence when the device development focus of the second world war-time research was replaced by peace-time research into the fundamental understanding of real semiconductors, the foundations had been laid for working devices based on elemental semiconductors to be realised.

1.2 The Silicon Age

1.2.1 The Transistor and Early Semiconductor Materials Development

As its name suggests, electronics is about the control of electrons to produce useful properties; electronic materials are the media in which this manipulation takes place. Exactly fifty years after J. J. Thompson had discovered the electron in 1897, mankind's ability to control them underwent a revolution due to the discovery of the transistor effect. It could be said that the world began to change in the final couple of weeks of 1947 when John Bardeen and Walter Brattain used germanium to build and demonstrate the first "semi-conductor

triode" (a device later to be named the point contact transistor to reflect its transresistive properties). This success at Bell Laboratories was obtained within just a few years of the post-war establishment of a research group led by William Shockley focussing on the understanding of semiconducting materials. It was to earn Brattain, Bardeen and Shockley the 1956 Nobel Prize for Physics.

The first point-contact transistor was based around three contacts onto an n-doped germanium block: when a small current passed between the "base" and "emitter", an amplified current would flow between the "collector" and "emitter" [1.6]. The emitter and collector

contacts needed to be located very close to one another (50–250 μm) and this was achieved by evaporating gold onto the corner of a plastic triangle, cutting the film with a razor blade and touching this onto the germanium – the two isolated strips of gold serving as the two contacts [1.7]. At about one centimetre in height, based on relatively impure polycrystalline germanium and adopting a different principle of operation, the device bears barely any resemblance to today's integrated circuit electronics components. Nonetheless it was the first implementation of a solid-state device capable of modulating (necessary for signal amplification in communications) and switching (needed for logic operations in computing) an electric current. In a world whose electronics were delivered by the thermionic vacuum tube, the transistor was immediately identified as a component which could be “employed as an amplifier, oscillator, and for other purposes for which vacuum tubes are ordinarily used” [1.6].

In spite of this, after the public announcement of the invention at the end of June 1948 the response of both the popular and technical press was somewhat muted. It was after all still “little more than a laboratory curiosity” [1.8] and ultimately point-contact transistors were never suited to mass production. The individual devices differed significantly in characteristics, the noise levels in amplification were high and they were rapidly to be superseded by improved transistor types.

A huge range of transistor designs have been introduced from the late 1940s through to today. These successive generations either drew upon, or served as a catalyst for, a range of innovations in semiconductor materials processing and understanding. There are many fascinating differences in device design but from a materials science point of view the three most striking differences between the first point contact transistor and the majority of electronics in use today are the choice of semiconductor, the purity of this material and its crystalline quality. Many of the key electronic materials technologies of today derive from the developments in these fields in the very early years of the post-war semiconductor industry.

Both germanium and silicon had been produced with increasing purity throughout the 1940s [1.9]. Principally because of germanium's lower melting temperature (937 °C compared with 1415 °C) and lower chemical reactivity its preparation had always proved easier and was therefore favoured for the early device manufacture such as the first transistor. However, the properties of silicon make it a much more attractive choice for solid state devices. While germanium is expensive and

rare, silicon is, after oxygen, the second most abundant element. Silicon has a higher breakdown field and a greater power handling ability; its semiconductor band gap (1.1 eV at 300 K; Fig. 1.1) is substantially higher than germanium's (0.7 eV) so silicon devices are able to operate over a greater range of temperatures without intrinsic conductivity interfering with performance.

The two materials competed with one another in device applications until the introduction of novel doping techniques in the mid-1950s. Previously p- and n-doping had been achieved by the addition of dopant impurities to the semiconductor melt during solidification. A far more flexible technique involved the diffusion of dopants from the vapour phase into the solid semiconductor surface [1.10]. It became possible to dope with a degree of two-dimensional precision when it was discovered that silicon's oxide served as an effective mask to dopant atoms and that a photoresist could be used to control the etching away of the oxide [1.11, 12]. Successful diffusion masks could not be found for germanium and it was soon abandoned for mainstream device manufacture. Dopant diffusion of this sort has since been superseded by the implantation of high-energy ions which affords greater control and versatility.

Shockley was always aware that the material of the late 1940s was nothing like pure enough to make reliable high performance commercial devices. Quantum mechanics suggested that to make a high quality transistor out of the materials it was necessary to reduce the impurity level to about one part in 10^{10} . This was a far higher degree of purity than existed in any known material. However, William Pfann, who worked at Bell Laboratories, came up with the solution. He invented a technique called zone-refining to solve this problem, and showed that repeated zone refining of germanium and silicon reduced the impurities to the level required. The work of *Pfann* is not widely known but was a critical piece of materials science which enabled the practical development of the transistor [1.13, 14].

At a similar time great progress was being made in reducing the crystalline defect density of semiconducting materials. Following initial hostility by some of the major researchers in the field it was rapidly accepted that transistor devices should adopt single crystalline material [1.15]. Extended single crystals of germanium several centimetres long and up to two centimetres in diameter [1.15, 16] and later similar silicon crystals [1.17] were produced using the *Czochralski* technique of pulling a seed crystal from a high purity melt [1.18]. The majority of material in use today is derived from this route. To produce silicon with

the very lowest impurity concentration, an alternative method called float zoning was developed where a polycrystalline rod was converted to a single crystal by the passage of a surface tension confined molten zone along its length [1.19–21]. No crucible is required in the process so there are fewer sources of impurity contamination. Float zoning is used to manufacture some of the purest material in current use [1.22]. The early Czochralski material contained dislocation densities of 10^5 – 10^6 cm⁻² but by the start of the 1960s dislocation free material was obtained [1.23–26]. Initially most wafers were on the silicon (111) plane, which was easiest to grow, cut and polish [1.27]. For field-effect devices, which are discussed below, use of the (100) plane was found to offer preferable properties so this was introduced in the same decade. The impurity concentration in dislocation-free silicon has been continually reduced up to the present day and wafer diameters have increased almost linearly (though accelerating somewhat in recent years) from about 10 mm in the early 1960s to the “dinner plate” 300 mm today [1.22]. These improvements represent one of the major achievements in semiconductor materials growth and processing.

A series of generations of transistors followed in rapid succession after Brattain and Bardeen’s first triumph. Here we only mention a few of the major designs whose production have traits in common with technology today. Early in 1948 *Shockley* developed a detailed formulation of the theory of p–n junctions that concluded with the conception of the junction transistor [1.28, 29]. This involved a thin n-doped base layer sandwiched between p-doped emitter and collector layers (or vice versa). This p–n–p (n–p–n) structure is the simplest form of the bipolar transistor (so-called because of its use of both positive and negative charge carriers), a technology which remains important in analogue and high-speed digital integrated circuits today. In April 1950, by successively adding arsenic and gallium (n- and p-type dopants respectively) impurities to the melt, n–p–n junction structures with the required p-layer thickness (≈ 25 μ m) were formed from single crystal germanium. When contacts were applied to the three regions the devices behaved much as expected from *Shockley’s* theory [1.28, 30]. Growth of junction transistors in silicon occurred shortly afterwards and they entered production by Texas Instruments in 1954 [1.15].

By the later years of the 1950s, the diffusion doping technique was used to improve the transistor’s speed response by reducing the thickness of the base layer in the

diffused base transistor [1.31]. This began the trend of manufacturing a device in situ on a substrate material so in a sense it was the foundation for all subsequent micro-electronic structures. Soon afterwards, epitaxial growth techniques were introduced [what would today be described as vapour phase epitaxy (VPE)] which have since become central to both silicon and compound semiconductor technology. Gas phase precursors were reacted to produce very high quality and lightly doped crystalline silicon on heavily doped substrate wafers to form epitaxial diffused transistors. Since the collector contact was made through the thickness of the wafer, the use of highly doped (low resistance) wafers reduced the series resistance and therefore increased the frequency response [1.32].

For some years the highest performance devices were manufactured using the so-called “mesa” process where the emitter and diffused base were raised above the collector using selective etching of the silicon [1.28]. The planar process (which is still at the heart of device production today) was subsequently developed, in which the p–n junctions were all formed inside the substrate using oxide masking and diffusion from the surface. This resulted in a flat surface to which contacts could be made using a patterned evaporated film [1.33]. This processing technique was combined with some exciting thoughts at the end of the 1950s and led to the application of transistor devices and other components in a way which was to transform the world: the integrated circuit.

1.2.2 The Integrated Circuit

With the benefit of hindsight, the integrated circuit concept is quite simple. The problem faced by the electronics industry in the 1950s was the increasing difficulty of physically fitting into a small device all of the discrete electronic components (transistors, diodes, resistors and capacitors), and then connecting them together. It was clear that this problem would eventually limit the complexity, reliability and speed of circuits which could be created. Transistors and diodes were manufactured from semiconductors but resistors and capacitors were best formed from alternative materials. Even though they would not deliver the levels of performance achievable from the traditional materials, functioning capacitors and resistors *could* be manufactured from semiconductors so, in principle, all of the components of a circuit could be prepared on a single block of semiconducting material. This reasoning had been proposed by Englishman G.W.A. Dummer at

a conference in 1952 [1.34] but small-scale attempts to realise circuits had failed, largely because they were based on connecting together layers in grown-junction transistors [1.35]. In 1958, however, Jack Kilby successfully built a simple oscillator and “flip-flop” logic circuits from components formed in situ on a germanium block and interconnected to produce circuits. He received the Nobel Prize in 2000 for “his part in the invention of the integrated circuit”.

Kilby’s circuits were the first built on a single semiconductor block, but by far the majority of the circuit’s size was taken up by the wires connecting together the components. Robert Noyce developed a truly integrated circuit (IC) in the form that it was later to be manufactured. While Kilby had used the mesa technique with external wiring, Noyce applied the planer technique to form transistors on silicon and photolithographically defined gold or aluminium interconnects. This was more suited to batch processing in production and was necessary for circuits with large numbers of components.

Most integrated circuits manufactured today are based around a transistor technology distinct from the bipolar device used in the first chips but one still dating from the 1960s. In 1960 the first metal oxide semiconductor field effect transistor (MOSFET) was demonstrated [1.36]. In this device a “gate” was deposited onto a thin insulating oxide layer on the silicon. The application of a voltage to the gate resulted in an inversion layer in the silicon below the oxide thereby modifying the conducting channel between “source” and “drain” contacts. This structure was a p-MOS device (current transfer between the collector and emitter was by hole conduction) grown on (111) silicon using an aluminium gate. Earlier attempts at such a device had failed because of trapped impurities and charges in the gate oxide – this new structure had reduced the density of these to below tolerable levels but the device still could not compete with the bipolar transistors of its time [1.27]. By 1967, however, (100) silicon (which offered lower densities of states at the Si/SiO₂ interface) was used together with a polycrystalline silicon gate to construct a more effective and more easily processed device with advantages over the bipolar transistor. In the early 1970s the n-MOS device, which was even less tolerant to the positive gate oxide charges, was realised thanks to much improved cleanliness in the production environment. With conduction occurring by the transfer of electrons rather than holes these were capable of faster operation than similar p-MOS structures (the mobility of electrons in silicon is about three times that of

holes). By the 1980s, these two devices were combined in the complementary MOS (CMOS) device which afforded much lower power consumption and simplified circuit design [1.37]. This remains the principal structure used in microelectronics today. Of course now it is much smaller and significantly faster thanks to a range of further advances. These include improved control of the doping and oxidation of silicon and developments in optical lithographic techniques [1.37]. The minimum dimension of components which can be lithographically patterned on an integrated circuit is ultimately limited by the wavelength of radiation used in the process and this has continually been decreased over the past few decades. In the late 1980s wavelengths of 365 nm were employed; by the late 1990s 248 nm were common and today 193 nm is being used. Research into extreme ultra-violet lithography at 13.5 nm may see this being adopted within the next decade enabling feature sizes perhaps as low as 25 nm.

Though ICs implementing CMOS devices are the foundations for computing, silicon-based bipolar transistors maintain a strong market position today in radio frequency applications. In particular germanium is making something of a comeback as a constituent of the latest generation of SiGe bipolar devices [1.38]. The combination of the two forms of technology on a single chip (BiCMOS) offers the potential for computing and communications to be integrated together in the wireless devices of the coming decade.

The development in complexity and performance of silicon devices, largely due to materials science progress, is unparalleled in the history of technology. Never before could improvements be measured in terms of a logarithmic scale for such a sustained period. This is often seen as the embodiment of “Moore’s Law”. Noting a doubling of the number of components fitted onto integrated circuits each year between 1959 and 1965, *Moore* predicted that this rate of progress would continue until at least ten years later [1.39]. From the early 1970s, a modified prediction of doubling the number of components every couple of years has been sustained to the current day. Since the goals for innovation have often been defined assuming the continuation of the trend, it should perhaps be viewed more as a self-fulfilling prophecy. A huge variety of statistics relating to the silicon microelectronics industry follow a logarithmically scaled improvement from the late 1960s to the current day: the number of transistors shipped per year (increasing); average transistor price (decreasing); and number of transistors on a single chip (increasing) are examples [1.40]. The final member of this list is

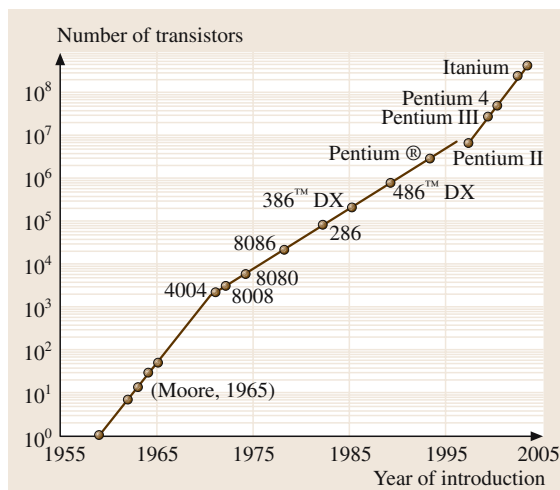


Fig. 1.2 The realisation of Moore's law by commercial Intel® processors: The logarithmic increase in the number of transistors in each processor chip

plotted in Fig. 1.2. A graphic example of the continuous scaling down of IC component dimensions is the reduction in the thickness of the MOSFET gate insulator. Figure 1.3 shows high-resolution transmission electron microscopy images of typical commercial devices since the mid-1990s: improved device performance has pre-

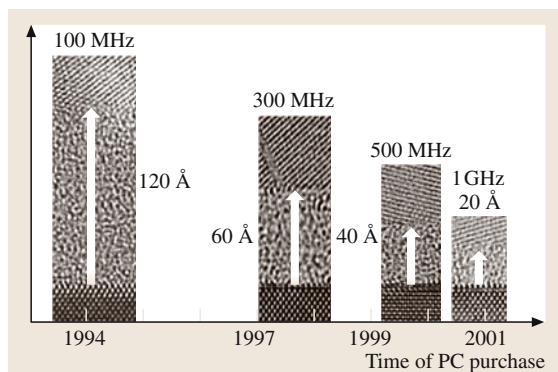


Fig. 1.3 The relentless march to zero thickness. HRTEM images of SiO_2 gate oxide thickness used in personal computers (After [1.41] with permission from the Materials Research Society). (Courtesy of Prof D.A. Muller)

dominately been achieved by reducing all dimensions of the IC components. With state-of-the-art gates now approaching the sub-nanometre thicknesses at which SiO_2 breaks down as an insulator, it is clear that achieving device performance by scaling alone cannot be sustained for much longer. Overcoming this limit, perhaps through the substitution of SiO_2 for a material with a higher dielectric constant, is one of the major challenges in silicon IC research today.

1.3 The Compound Semiconductors

It has been said that silicon is to electronics what steel is to mechanical engineering [1.42]. Steel is very effectively used for most of the world's construction but there are some tasks which it is incapable of performing and others for which alternative structural materials are better suited. In the same way there are some crucial applications – such as optoelectronics and very high speed electronics – that silicon cannot usually deliver but which a wide range of compound semiconductors are better equipped to perform.

Silicon's band gap is indirect (an electron-hole recombination across the band gap must be accompanied by an interaction with a phonon in the lattice) which severely limits the potential efficiency of light emission from the material. Many of the important compound semiconductors, such as the alloys $\text{Al}_x\text{Ga}_{1-x}\text{As}$, $\text{In}_x\text{Ga}_{1-x}\text{N}$, $\text{Al}_y\text{Ga}_{1-y}\text{N}$, and $\text{Al}_x\text{In}_y\text{Ga}_{1-x-y}\text{P}$ exhibit direct band gaps (no phonon interaction is required) so can efficiently emit brilliant light in light emitting diodes (LEDs) and laser diodes. Furthermore, in these alloy

systems, where the band gap can be adjusted by changing the composition, there is a means of selecting the energy released when an electron and hole recombine across the gap and therefore controlling the wavelength of the photons emitted. From the $\text{Al}_x\text{In}_y\text{Ga}_{1-x-y}\text{As}$ and $\text{Al}_y\text{In}_x\text{Ga}_{1-x-y}\text{N}$ alloy systems there is, in principle, a continuous range of direct band gaps from deep in the infra red (InAs; $\lambda = 3.5 \mu\text{m}$) to far into the ultraviolet (AlN; $\lambda = 200 \text{ nm}$). The semiconductor band gaps of these materials and the corresponding photon wavelengths are put into context with the visible spectrum in Fig. 1.1.

Compounds are also very useful in high speed electronics applications. One of the determining factors in the speed of a transistor is the velocity of the charge carriers in the semiconductor. In GaAs the electron drift velocity is much higher than in silicon so its transistors are able to operate at significantly higher frequencies. The electron velocity in InAs is higher still. Furthermore, in the same way that silicon was preferred to

germanium, devices manufactured using semiconductors such as GaN, which have much wider band gaps than silicon (3.4 eV compared with 1.1 eV), are capable of operating in much higher temperature environments.

Aside from these advantageous properties of compound semiconductors, the use of different alloy compositions, or totally different semiconductors, in a single device introduces entirely new possibilities. In silicon, most device action is achieved by little more than careful control of dopant impurity concentrations. In structures containing thin layers of semiconductors with different band gaps (heterostructures) there is the potential to control more fundamental parameters such as the band gap width, mobilities and effective masses of the carriers [1.42]. In these structures, important new features become available which can be used by the device designer to tailor specific desired properties. Hebert Kroemer and Zhores Alferov shared the Nobel prize in 2000 “for developing semiconductor heterostructures used in high-speed- and opto-electronics”.

We will mainly consider the compounds formed between elements in Group III of the periodic table and those in Group V (the III–V semiconductors); principally those based around GaAs and InP which were developed over much of the last forty years, and GaN and its related alloys which have been most heavily studied only during the last decade. Other families are given less attention here though they also have important applications (for example the II–VI materials in optoelectronic applications). It can be hazardous to try and consider the “compound semiconductors” as a single subject. Though lessons can be learnt from the materials science of one of the compounds and transferred to another, each material is unique and must be considered on its own (that is, of course, the purpose of the specialised chapters which follow in this handbook!).

It is worth repeating that the power of the compound semiconductors lies in their use as the constituent layers in heterostructures. The principal contribution from chemistry and materials science to enable successful devices has been in the manufacture of high-quality bulk single crystal substrates and the creation of techniques to reliably and accurately produce real layered structures on these substrates from the plans drawn up by a device theorist. In contrast to silicon, the compound semiconductors include volatile components so encapsulation has been required for the synthesis of low-defect InP and GaAs substrates such as in the liquid encapsulated Czochralski technique [1.43, 44]. The size and crystalline quality of these substrates lag some way behind those available in silicon. Crucial to the commerciali-

sation of electronic and optoelectronic heterostructures were the improvements over the last few decades in the control of epitaxial growth available to the crystal grower. The first successful heterostructures were manufactured using deposition onto a substrate from the liquid phase (liquid phase epitaxy; **LPE**) – “a beautifully simple technology but with severe limitations” [1.42]. However, the real heterostructure revolution had to wait for the 1970s and the introduction of molecular beam epitaxy (**MBE**) and metalorganic chemical vapour deposition (**MOCVD**) – also known as metalorganic vapour phase epitaxy (**MOVPE**) provided that the deposition is epitaxial.

MBE growth occurs in an ultra-high vacuum with the atoms emitted from effusion cells forming “beams” which impinge upon, and form compounds at the substrate surface. It derives from pioneering work at the start of the 1970s [1.45]. **MOCVD** relies on chemical reactions occurring on the substrate involving metalorganic vapour phase precursors and also stems from initial work at this time [1.46]. In contrast to **LPE**, these two techniques permit the combination of a wide range of different semiconductors in a single structure and offer a high degree of control over the local composition, in some cases on an atomic layer scale. The successful heterostructure devices of the late 1970s and 1980s would not have been achievable without these two tools and they still dominate III–V device production and research today.

1.3.1 High Speed Electronics

The advantages of the III–V materials over silicon for use in transistors capable of operating at high frequencies were identified early in the semiconductor revolution [1.47]. Shockley’s first patent for p-n junction transistors had included the proposal to use a wide-gap emitter layer to improve performance and in the 1950s *Kroemer* presented a theoretical design for a heterostructure transistor [1.48]. Some years later the structure of a GaAs metal semiconductor **FET** (**MESFET**) was proposed and realised soon afterwards [1.49, 50]. In these devices a Schottky barrier surface potential was used to modulate the conductivity of the GaAs channel. One of the earliest applications of the III–V’s was as low noise amplifiers in microwave receivers which offered substantial improvements relative to the silicon bipolar transistors of the time. The devices were later used to demonstrate sub-nanosecond switching in monolithic digital **ICs** [1.51]. Today they form the core of the highest speed digital circuits and are used in

high speed electronics in microwave radar systems and wireless communications which incorporate monolithic integrated circuits.

For at least 30 years there have been repeated attempts to replicate the **MOSFET**, the dominant transistor form in silicon **ICs**, on GaAs material. These attempts have been frustrated by the difficulty of reproducibly forming a high quality stoichiometric oxide on GaAs. In direct analogy with the initial failure of constructing working **n-MOSFETs** on silicon, the GaAs devices have consistently been inoperable because of poor quality gate oxides with a high density of surface states at the GaAs-insulator interface [1.44]. One of the research efforts focussed on realising this device was, however, to be diverted and resulted in the discovery of probably the most important III–V electronic device: the high electron mobility transistor (**HEMT**).

The background to this invention lies in the beautiful concept of modulation doping of semiconductors which was first demonstrated in 1978 [1.52]. One of the tenets of undergraduate semiconductor courses is the demonstration that as the dopant density in a semiconductor increases, the mobility of the carriers is reduced because the carriers are scattered more by the ionised dopants. It was found that in a multilayer of repeating **n-AlGaAs** layers and undoped GaAs layers, the electrons supplied by donor atoms in the AlGaAs moved into the adjacent potential wells of the lower-band gap GaAs layers. In the GaAs these suffered from substantially less ionised impurity scattering and therefore demonstrated enhanced mobility.

While working in a group attempting to create GaAs **MOSFETs** (and seemingly despairing at the task! [1.53]), *Mimura* heard of these results and conceived of a field effect transistor where the conducting channel exploited the high mobility associated with a modulation doped structure. In essence, a doped AlGaAs layer was formed above the undoped GaAs channel of the transistor. Donated carriers gathered in the GaAs immediately below the interface where they did not suffer from as much ionised impurity scattering and so their mobility would approach that of an ultra-pure bulk semiconductor. The current was conducted from the source to the drain by these high mobility carriers and so the devices were able to operate in higher frequency applications [1.53]. Realisation of the structure required a very abrupt interface between the GaAs and AlGaAs and was considered beyond the capability of **MOCVD** of the time [1.53]. However, following the advances made in **MBE** procedures during the 1970s the structure was achieved by that technique within a few months of

the original conception [1.53, 54]. The first operational **HEMT** chips were produced on 24th December 1980: by pleasing coincidence this was the anniversary of Brattain and Bardeen's demonstration of their point contact resistor to the management of Bell labs in 1947! Structures based on the same principle as Mimura's device were realised in France very shortly afterwards [1.55].

The commercialisation of the **HEMT** became significant in the late 1980s thanks to broadcasting satellite receivers. The improved performance of the devices compared with the existing technology allowed the satellite parabolic dish size to be reduced by at least a factor of two. Structures similar to these have since played a crucial role in the massive expansion in mobile telephones.

The evolution in **HEMT** structures since the early 1980s is a fine example of how fundamental compound semiconductor properties have been exploited as the technology has become available to realise new device designs. The electron mobility in InAs is much higher than in GaAs and rises as the indium content in $\text{In}_x\text{Ga}_{1-x}\text{As}$ is increased [1.56]. The introduction of an InGaAs (as opposed to GaAs) channel to the **HEMT** structure resulted both in increased electron mobility and a higher density of carriers gathering from the doped AlGaAs layer (because of the larger difference in energy between the conduction band minima of InGaAs and AlGaAs than between GaAs and AlGaAs). This so called pseudomorphic **HEMT** (**pHEMT**) demonstrates state of the art power performance at microwave and millimetre wave frequencies [1.43]. The indium content and thickness of the channel is limited by the lattice mismatch with the GaAs (Fig. 1.1). If either is increased too much then misfit dislocations are formed within the channel. The restriction is reduced by growing lattice matched structures on InP, rather than GaAs, substrates. $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ are both lattice matched to the InP (Fig. 1.1) and their conduction band minimum energies are well separated so that in the InGaAs below the interface between the two compounds a high density of electrons with a very high mobility is formed. Compared with the **pHEMTs** these InP based **HEMTs** exhibit significant improvements, have been shown to exhibit gain at over 200 GHz and are established as the leading transistor for millimetre-wave low noise applications such as radar [1.43].

1.3.2 Light Emitting Devices

LEDs and laser diodes exploit the direct band gap semiconductors to efficiently convert an electric current into

photons of light. Work on light emission from semiconductor diodes was carried out in the early decades of the twentieth century [1.57] but the start of the modern era of semiconductor optoelectronics traces from the demonstration of LED behaviour and lasing from p-n junctions in GaAs [1.58, 59] and $\text{GaAs}_{1-x}\text{P}_x$ [1.60]. The efficiency of these LEDs was low and the lasers had large threshold currents and only operated at low temperatures. A year later, in 1963, *Kroemer* and *Alferov* independently proposed the concept of the double heterostructure (DH) laser [1.61, 62]. In the DH device, a narrow band gap material was to be sandwiched between layers with a wider gap so that there would be some degree of confinement of carriers in the “active layer”. By the end of the decade DH devices had been constructed which exhibited continuous lasing at room temperature [1.63, 64]. Alferov’s laser was grown by LPE on a GaAs substrate with a $0.5\text{ }\mu\text{m}$ GaAs active layer confined between $3\text{ }\mu\text{m}$ of $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$ on either side. The launch of the Compact Disc in 1982 saw this type of device, or at least its offspring, becoming taken for granted in the households of the world.

One of the major challenges in materials selection for heterostructure manufacture has always been avoiding the formation of misfit dislocations to relieve the strain associated with lattice parameter mismatch between the layers. $\text{Al}_x\text{Ga}_{1-x}\text{As}$ exhibits a direct band gap for $x < 0.45$ and the early success and sustained dominance of the AlGaAs/GaAs system derives significantly from the very close coincidence of the AlAs and GaAs lattice parameters ($5.661\text{ }\text{\AA}$ and $5.653\text{ }\text{\AA}$ – see Fig. 1.1). This allows relatively thick layers of AlGaAs with reasonably high aluminium content to be grown lattice matched onto GaAs substrates with no misfit dislocation formation. The use of the quaternary alloy solid solution $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ was also suggested in 1970 [1.63] to offer the independent control of lattice parameters and band gaps. Quaternaries based on three Group III elements have since proved very powerful tools for lattice matching within heterostructures. $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ was found to be almost perfectly lattice matched to GaAs and additionally have a very similar thermal expansion coefficient (which is important to avoid strain evolution when cooling after growth of heteroepitaxial layers at high temperatures). By varying x in this compound, direct band gaps corresponding to light between red and green could be created [1.65]. Lasers based on this alloy grown by MOCVD are a common choice for the red wavelengths (650 nm) used in DVD reading.

Obtaining lattice matching is not so crucial for layers thinner than the critical thickness for dislocation

production and can be less of an issue these days because of probably the most important development in the history of optoelectronic devices: the introduction of the quantum well. In some ways a quantum well structure is an evolution of the double heterostructure but with a very much thinner active layer. It is the chosen design for most solid state light emitting devices today. With the accurate control available from MBE or MOCVD, and following from some early work on superlattices [1.66], very thin layers of carefully controlled composition could be deposited within heterostructure superlattice stacks. It became possible to grow GaAs layers much less than 10 nm thick within AlGaAs–GaAs heterostructures. The carriers in the GaAs were found to exhibit quantum mechanical confinement within the one dimensional potential well [1.67, 68]. Lasing from GaAs/ $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$ quantum wells was reported the following year, in 1975, [1.69] but it was a few years before the performance matched that achievable from DH lasers of the time [1.70] and the quantum well laser was further advanced to significantly outperform the competition by researchers in the 1980s [1.71].

The introduction of heterostructures with layer thicknesses on the nanometre scale represents the final stage in scaling down of these devices. Similarly Brattain and Bardeen’s centimetre-sized transistor has evolved into today’s microprocessors with sub-micron FETs whose gate oxide thicknesses are measured in Angstroms. Throughout this evolution, materials characterisation techniques have contributed heavily to the progress in our understanding of electronic materials and deserve a brief detour here. As the dimensions have been reduced over the decades, the cross-sectional images of device structures published in the literature have changed from a period where optical microscopy techniques were sufficient [1.31] to a time when scanning electron microscopy (SEM) images were used [1.70] and to today’s high resolution transmission electron microscopy (TEM) analysis of ultra-thin layers (e.g. Fig. 1.3). For each new material family, understanding of defects and measurement of their densities (e.g. by TEM and X-ray topography) have contributed to improvements in quality. Huge improvements in X-ray optics have seen high-resolution X-ray diffraction techniques develop to become a cornerstone of heterostructure research and production quality control [1.72]. Scanning-probe techniques such as scanning tunnelling and atomic force microscopy have become crucial to the understanding of MBE and MOCVD growth. Chemically sensitive techniques such as secondary ion mass spectroscopy and Rutherford

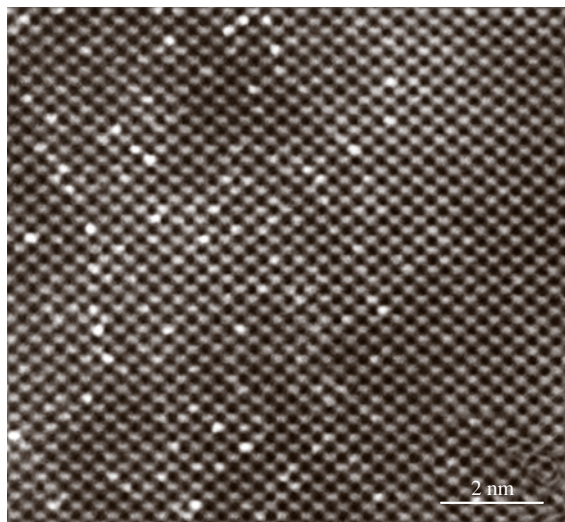


Fig. 1.4 Annular dark field-scanning transmission electron microscope (ADF-STEM) image of Sb-doped Si. The undoped region (*right*) shows atomic columns of uniform intensity. The brightest columns in the doped region (*left*) contain at least one Sb atom. The image is smoothed and background subtracted (After [1.76], with permission Elsevier Amsterdam). (Courtesy of Prof D.A. Muller)

backscattering have improved to provide information on doping concentrations and compositions in layered structures with excellent depth resolution. Meanwhile recently developed techniques such as energy-filtered TEM [1.73] afford chemical information at extremely high spatial resolutions. The characterisation of doping properties is also coming of age with more quantitative measurement of dopant contrast in the SEM [1.74], analysis of biased junctions in situ in the TEM [1.75] and the recent exciting demonstration of imaging of single impurity atoms in a silicon sample using scanning TEM, Fig. 1.4. The materials characterisation process remains a very important component of electronic materials research.

Two commercial-product oriented aims dominate semiconductor laser research: the production of more effective emitters of infra-red wavelengths for transmission of data along optic fibres; and the realisation of shorter wavelength devices for reading optical storage media. In the first of these fields devices based on InP have proven to be extremely effective because of its fortuitous lattice parameter match with other III–V alloys which have band gaps corresponding to the low-absorption “windows” in optic fibres. While remaining lattice matched to InP, the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ quater-

nary can exhibit band gaps corresponding to infra-red wavelengths of 1.3 μm and 1.55 μm at which conventional optic fibres absorb the least of the radiation (the absolute minimum is for 1.55 μm). Room temperature continuous lasing of 1.1 μm radiation was demonstrated from the material in 1976 [1.77] and InP based lasers and photodiodes have played a key role in the optical communications industry since the 1980s [1.43].

We have already mentioned the AlGaAs infra-red ($\lambda = 780\text{ nm}$) emitters used to read compact discs and the AlGaInP red ($\lambda = 650\text{ nm}$) devices in DVD readers (see Fig. 1.1). As shorter wavelength lasers have become available the optical disc’s surface pits (through which bits of data are stored) could be made smaller and the storage density increased. Though wide band gap II–VI compounds, principally ZnSe, have been researched for many decades for their potential in green and blue wavelengths, laser operation in this part of the visible spectrum proved difficult to realise [1.78]. In the early 1990s, following improvements in the p-doping of ZnSe, a blue-green laser was demonstrated [1.79] but such devices remain prone to rapid deterioration during operation and tend to have lifetimes measured in, at most, minutes. However, also in the early years of the 1990s, a revolution began in wide band gap semiconductors which is ongoing today: the exploitation of GaN and its related alloys $\text{In}_x\text{Ga}_{1-x}\text{N}$ and $\text{Al}_y\text{Ga}_{1-y}\text{N}$. These materials represent the future for optoelectronics over a wide range of previously inaccessible wavelengths and the next generation of optical storage, the “Blu-ray” disc, will be read using an InGaN blue-violet laser ($\lambda = 405\text{ nm}$).

1.3.3 The III-Nitrides

The relevance of the $\text{In}_x\text{Ga}_{1-x}\text{N}$ alloy for light emitting devices is clear from Fig. 1.1. The InN and GaN direct band gaps correspond to wavelengths straddling the visible spectrum and the alloy potentially offers access to all points in-between. The early commercially successful blue light emitters were marketed by Nichia Chemical Industries following the research work of *Nakamura* who demonstrated the first InGaN DH LEDs [1.80] and blue InGaN quantum well LEDs and laser diodes soon after [1.81]. Since this time the global research interest in the GaN material family has expanded rapidly and the competing technology (SiC and ZnSe for blue LEDs and lasers respectively) has largely been replaced.

The development of the III-nitride materials has much in common with the early research of other III–V systems. For example MOCVD and MBE technology

could be adapted for the nitride systems (the former has to date been more suited for creating optical devices) and one of the obstacles limiting early device development was achieving sufficiently high p-type doping. However, in some ways they are rather different from the other compound semiconductors. It is important to realise that while all III–Vs mentioned previously share the same cubic crystallographic structure, the nitrides most readily form in a hexagonal allotrope. Most significant in terms of device development over the past decade has been the difficulty in obtaining bulk GaN substrates. Due to the very high pressures necessary to synthesise the compound only very small pieces of bulk GaN have been produced and though they have been used to form functioning lasers [1.82] they remain unsuited as yet to commercial device production. There has consequently been a reliance on heavily lattice mismatched heteroepitaxial growth.

Many materials have been used as substrates for GaN growth. These include SiC, which has one of the lowest lattice mismatches with the nitride material and would be more widely used if it was less expensive, and silicon, which has considerable potential as a substrate if problems associated with cracking during cooling from the growth temperature can be overcome (there is a large difference between the thermal expansion coefficients of GaN and silicon). The dominant choice, however, remains sapphire (α -Al₂O₃) which itself is by no means ideal: it is electrically insulating (so electrical contacts cannot be made to the device through the substrate material) and, most significantly, has a lattice mismatch of $\approx 16\%$ with the GaN [1.83]. This mismatch is relieved by the formation of misfit dislocations which give rise to dislocations threading through the GaN into the active layers (e.g. InGaN quantum wells) of the devices. The key discovery for reducing the defect densities to tolerable levels was the use of nucleation layers at the interface with the sapphire [1.83] but densities of $\approx 10^9 \text{ cm}^{-2}$ remain typical. More recently epitaxial lateral overgrowth (ELOG) techniques have allowed the dislocation densities to be reduced to $\approx 10^6 \text{ cm}^{-2}$ in local regions [1.84]. However, perhaps the most interesting aspect of GaN-based optoelectronic devices is that they emit light so efficiently in spite of dislocation density orders of magnitude greater than those tolerated in conventional semiconductors.

Even though InGaN based light emitters have been commercially available for several years, the precise mechanism of luminescence from the alloy is still not fully understood. Having so far discussed the evolution of semiconductors with the benefit of hindsight we

can now for a moment consider an unresolved issue which, no doubt, will be solved in the coming years. It is widely believed that the tolerance of InGaN optoelectronic devices to high densities of defects is caused by the presence of low-energy sites within the layers at which electrons and holes are localised. They are thus prevented from interacting with the dislocations at which they would recombine in a non-radiative manner. The origin of localisation remains a matter of debate. One popular explanation is that the InGaN alloy has a tendency for phase segregation [1.85] and indium-rich “clusters” form and cause the localisation. However, there is now evidence [1.86] that the results of some of the measurements used to detect the indium rich regions could be misleading so the clustering explanation is being re-assessed. InGaN remains a fascinating and mysterious alloy.

Solid-state lighting will be a huge market for III-nitride materials in the coming decades. LEDs are perfectly suited to coloured light applications: their monochromatic emission is very much more energy efficient than the doubly wasteful process of colour filtering power hungry filament white light bulbs. InGaN LEDs are now the device of choice for green traffic signals worldwide and offer significant environmental benefits in the process. In principle there is also the opportunity to create white light sources for the home which are more efficient than the tungsten filament light bulbs used today and a variety of promising schemes have been developed for converting the coloured output of III-nitride LEDs to white sources. These include the use of three colour (red, green and blue) structures and blue InGaN or ultraviolet AlGaIn based LEDs coated with a range of phosphor materials to generate a useful white spectrum [1.65]. In particular ultraviolet LEDs coated with a three-way phosphor (red, green and blue) can produce high quality white light that mimics sunlight in its visible spectrum. The main disadvantages preventing the widespread use of LEDs in white lighting are their high cost and the relatively low output powers from single devices but these obstacles are rapidly being overcome.

Many other applications for the III-nitrides are being investigated including the use of (Al,Ga,In)N solar cells which could offer more efficient conversion of light into electric current than silicon based devices [1.87]; the possibility of lasers and optical switches operating at the crucial $1.55 \mu\text{m}$ wavelength based on intersubband (between the discrete quantised energy levels of the wells) transitions in AlGaIn/GaN quantum well structures; and the use of the compact InGaN LEDs to fluoresce labelled cancerous cells and aid detection of

affected areas. The wide band gap is also very attractive for many electronic device applications—particularly in high-temperature, high-power applications. Exploiting its high thermal conductivity and insensitivity to high operating temperatures, GaN-based HEMTs may extend the power of mobile phone base stations and it

has even been suggested that GaN devices could be used as an alternative source of ignition in car engines. There is also, of course, the possibility of monolithically integrating electronic and optoelectronic action onto a single chip. GaN-related materials should prove to have a huge impact on the technology of the coming decades.

1.4 From Faraday to Today

So, we have come 170 years from Faraday's nineteenth century observation of semiconductivity to a world dominated by electronic materials and devices. The balance of power between the different semiconductor families is an unstable and unpredictable one. For example if inexpensive, high quality, low defect density GaN substrates can be produced this will revolutionise the applications of GaN-based materials in both optoelectronics and electronics. The only inevitable fact is that the electronics revolution will continue to be crucially dependent on electronic materials understanding and improvement. And while reading the more focussed chapters in this

book and concentrating on the very important minutiae of a particular field, it can be a good idea to remember the bigger picture and the fact that electronic materials are remarkable!

Cambridge, October 2003

Further References

In particular we recommend the transcripts of the Nobel Lectures given by Brattain, Bardeen, Shockley, Kilby, Kroemer and Alferov. Available in printed form as set out below and, for the latter three in video, at www.nobel.se.

References

- 1.1 M. Faraday: *Experimental Researches in Electricity, Vol I and II* (Dover, New York 1965) pp. 122–125 and pp. 426–427
- 1.2 W. Smith: *J. Soc. Telegraph Eng.* **2**, 31 (1873)
- 1.3 W. G. Adams: *Proc. R. Soc. London* **25**, 113 (1876)
- 1.4 A. H. Wilson: *Proc. R. Soc. London, Ser. A* **133**, 458 (1931)
- 1.5 A. H. Wilson: *Proc. R. Soc. London, Ser. A* **134**, 277 (1931)
- 1.6 J. Bardeen, W. H. Brattain: *Phys. Rev.*, **74**, 230 (1948)
- 1.7 J. Bardeen: *Nobel Lecture, Physics, 1942–1962* (Elsevier, Amsterdam 1956) www.nobel.se/physics/laureates/1956/shockley-lecture.html
- 1.8 E. Braun, S. MacDonald: *Revolution in Miniature*, 2 edn. (Cambridge Univ. Press, Cambridge 1982)
- 1.9 W. H. Brattain: *Nobel Lecture, Physics, 1942–1962* (Elsevier, Amsterdam 1956) www.nobel.se/physics/laureates/1956/brattain-lecture.html
- 1.10 C. S. Fuller: *Phys. Rev. (Ser 2)* **86**, 136 (1952)
- 1.11 I. Derick, C. J. Frosh: ,US Patent 2 802 760 (1955)
- 1.12 J. Andrus, W. L. Bond: ,US Patent 3 122 817 (1957)
- 1.13 W. G. Pfann: *Trans. Am. Inst. Mech. Eng.* **194**, 747 (1952)
- 1.14 W. G. Pfann: *Zone Melting* (Wiley, New York 1958)
- 1.15 G. K. Teal: *IEEE Trans. Electron. Dev.* **23**, 621 (1976)
- 1.16 G. K. Teal, J. B. Little: *Phys. Rev. (Ser 2)* **78**, 647 (1950)
- 1.17 G. K. Teal, E. Buehler: *Phys. Rev.* **87**, 190 (1952)
- 1.18 J. Czochralski: *Z. Phys. Chem.* **92**, 219 (1917)
- 1.19 P. H. Keck, M. J. E. Golay: *Phys. Rev.* **89**, 1297 (1953)
- 1.20 H. C. Theurer: ,US Patent 3 060 123 (1952)
- 1.21 H. C. Theurer: *Trans. Am. Inst. Mech. Eng.* **206**, 1316 (1956)
- 1.22 K. A. Jackson (Ed.): *Silicon Devices* (Wiley, Weinheim 1998)
- 1.23 W. C. Dash: *J. Appl. Phys.* **29**, 736 (1958)
- 1.24 W. C. Dash: *J. Appl. Phys.* **30**, 459 (1959)
- 1.25 W. C. Dash: *J. Appl. Phys.* **31**, 736 (1960)
- 1.26 G. Ziegler: *Z. Naturforsch.* **16a**, 219 (1961)
- 1.27 M. Grayson (Ed.): *Encyclopedia of Semiconductor Technology* (Wiley, New York 1984) p. 734
- 1.28 I. M. Ross: *Bell Labs Tech. J.* **2**(4), 3 (1997)
- 1.29 W. Shockley: *Bell Syst. Tech. J.* **28**(4), 435 (1949)
- 1.30 W. Shockley, M. Sparks, G. K. Teal: *Phys. Rev.* **83**, 151 (1951)
- 1.31 M. Tanenbaum, D. E. Thomas: *Bell Syst. Tech. J.* **35**, 1 (1956)
- 1.32 C. M. Melliar-Smith, D. E. Haggan, W. W. Troutman: *Bell Labs Tech. J.* **2**(4), 15 (1997)
- 1.33 J. A. Hoerni: *IRE Trans. Electron. Dev.* **7**, 178 (1960)
- 1.34 J. S. Kilby: *IEEE Trans. Electron. Dev.* **23**, 648 (1976)
- 1.35 J. S. Kilby: *Nobel Lectures in Physics: 1996–2000* (Imperial College Press, London 2000) www.nobel.se/physics/laureates/2000/kilby-lecture.html

- 1.36 D. Kahng, M. M. Atalla: *Silicon-Silicon Dioxide Field Induced Surface Devices* (Solid State Research Conference, Pittsburgh, Pennsylvania 1960)
- 1.37 J. T. Clemens: Bell Labs Tech. J. **2**(4), 76 (1997)
- 1.38 T. H. Ning: IEEE Trans. Electron. Dev. **48**, 2485 (2001)
- 1.39 G. E. Moore: Electronics **38**(8) (1965)
- 1.40 G. E. Moore: *International Solid State Circuits Conference* (2003)
- 1.41 F. H. Baumann: Mater. Res. Soc. Symp. **611**, C4.1.1–C4.1.12 (2000)
- 1.42 H. Kroemer: *Nobel Lectures in Physics: 1996–2000* (Imperial College Press, London 2000) www.nobel.se/physics/laureates/2000/kroemer-lecture.html
- 1.43 O. Wada, H. Hasegawa (Eds.): *InP-Based Materials and Devices* (Wiley, New York 1999)
- 1.44 C. Y. Chang, F. Kai: *GaAs High-Speed Devices* (Wiley, New York 1994)
- 1.45 A. Y. Cho: J. Vac. Sci. Technol., **8**, S31 (1971)
- 1.46 H. M. Manasevit: Appl. Phys. Lett. **12**, 156 (1968)
- 1.47 H. J. Welker: IEEE Trans. Electron. Dev. **23**, 664 (1976)
- 1.48 H. Kroemer: RCA Rev. **18**, 332 (1957)
- 1.49 C. A. Mead: Proc IEEE **54**, 307 (1966)
- 1.50 W. W. Hooper, W. I. Lehrer: Proc IEEE **55**, 1237 (1967)
- 1.51 R. van Tuyl, C. Liechti: IEEE Spectrum **14**(3), 41 (1977)
- 1.52 R. Dingle: Appl. Phys. Lett. **33**, 665 (1978)
- 1.53 T. Mimura: IEEE Trans. Microwave Theory Tech. **50**, 780 (2002)
- 1.54 T. Mimura: Jpn. J. Appl. Phys. **19**, L225 (1980)
- 1.55 D. Delagebeaudeuf: Electron. Lett. **16**, 667 (1980)
- 1.56 D. Chattopadhyay: J. Phys. C **14**, 891 (1981)
- 1.57 E. E. Loebner: IEEE Trans. Electron. Dev. **23**, 675 (1976)
- 1.58 R. N. Hall: Phys. Rev. Lett. **9**, 366 (1962)
- 1.59 M. I. Nathan: Appl. Phys. Lett. **1**, 62 (1962)
- 1.60 N. Holonyak: Appl. Phys. Lett. **1**, 82 (1962)
- 1.61 H. Kroemer: Proc. IEEE **51**, 1782 (1963)
- 1.62 Z. I. Alferov: *Nobel Lectures in Physics: 1996–2000* (Imperial College Press, London 2000) www.nobel.se/physics/laureates/2000/alferov-lecture.html
- 1.63 Z. I. Alferov: Fiz. Tekh. Poluprovodn. **4**, 1826 (1970) Translated in: Sov. Phys. – Semicond. **4**, 1573 (1971)
- 1.64 I. Hayashi: Appl. Phys. Lett. **17**, 109 (1970)
- 1.65 A. Žukauskas: *Introduction to Solid-State Lighting* (Wiley, New York 2002)
- 1.66 L. Esaki, R. Tsu: IBM J. Res. Dev. **14**, 61 (1970)
- 1.67 L. L. Chang: Appl. Phys. Lett. **24**, 593 (1974)
- 1.68 R. Dingle: Phys. Rev. Lett. **33**, 827 (1974)
- 1.69 J. P. van der Ziel: Appl. Phys. Lett. **26**, 463 (1975)
- 1.70 R. Dupuis: Appl. Phys. Lett. **32**, 295 (1978)
- 1.71 Z. I. Alferov: Semicond. **32**, 1 (1998)
- 1.72 D. K. Bowen, B. K. Tanner: *High Resolution X-ray Diffractometry and Topography* (Taylor Francis, London 1998)
- 1.73 L. Reimer, C. Deininger: *Energy-filtering Transmission Electron Microscopy* (Springer, Berlin, Heidelberg 1995)
- 1.74 C. Schönjahn: Appl. Phys. Lett. **83**, 293 (2003)
- 1.75 A. C. Twitchett: Phys. Rev. Lett. **88**, 238302 (2002)
- 1.76 P. M. Voyles: Ultramicrosc. **96**, 251–273 (2003)
- 1.77 J. J. Hsieh: Appl. Phys. Lett. **28**, 709 (1976)
- 1.78 H. E. Ruda (Ed.): *Widegap II–VI Compounds for Opto-electronic Applications* (Chapman Hall, London 1992)
- 1.79 M. A. Haase: Appl. Phys. Lett. **59**, 1272 (1991)
- 1.80 S. Nakamura: Appl. Phys. Lett. **64**, 1687 (1994)
- 1.81 S. Nakamura: Jpn. J. Appl. Phys. **35**, L74 (1996)
- 1.82 P. Prystawko: Phys. Status Solidi (a) **192**, 320 (2002)
- 1.83 S. Nakamura, G. Fasol: *The Blue Laser Diode* (Springer, Berlin, Heidelberg 1997)
- 1.84 B. Beaumont: Phys. Status Solidi (b) **227**, 1 (2001)
- 1.85 I. Ho, G. B. Stringfellow: Appl. Phys. Lett. **69**, 2701 (1996)
- 1.86 T. M. Smeeton: Appl. Phys. Lett. **83**, 5419 (2003)
- 1.87 J. Hogan: New Scientist, 24 (7th December 2002)

Part A Fundamentals

Part A Fundamental Properties

2 **Electrical Conduction in Metals and Semiconductors**

Safa Kasap, Saskatoon, Canada
Cyril Koughia, Saskatoon, Canada
Harry Ruda, Toronto, Canada
Robert Johanson, Saskatoon, Canada

3 **Optical Properties of Electronic Materials: Fundamentals and Characterization**

Safa Kasap, Saskatoon, Canada
Cyril Koughia, Saskatoon, Canada
Jai Singh, Darwin, Australia
Harry Ruda, Toronto, Canada
Stephen K. O'Leary, Regina, Canada

4 **Magnetic Properties of Electronic Materials**

Charbel Tannous, Brest Cedex, France
Jacek Gieraltowski, 29285 Brest Cedex, France

5 **Defects in Monocrystalline Silicon**

Wilfried von Ammon, Burghausen, Germany

6 **Diffusion in Semiconductors**

Derek Shaw, Hull, UK

7 **Photoconductivity in Materials Research**

Monica Brinza, Leuven, Belgium
Jan Willekens, Leuven, Belgium
Mohammed L. Benkhedir, Leuven, Belgium
Guy J. Adriaenssens, Leuven, Belgium

8 **Electronic Properties of Semiconductor Interfaces**

Winfried Mönch, Duisburg, Germany

9 **Charge Transport in Disordered Materials**

Sergei Baranovskii, Marburg, Germany
Oleg Rubel, Marburg, Germany

10 **Dielectric Response**

Leonard Dissado, Leicester, UK

11 **Ionic Conduction and Applications**

Harry L. Tuller, Cambridge, USA

Electrical Con

2. Electrical Conduction in Metals and Semiconductors

Electrical transport through materials is a large and complex field, and in this chapter we cover only a few aspects that are relevant to practical applications. We start with a review of the semi-classical approach that leads to the concepts of drift velocity, mobility and conductivity, from which Matthiessen's Rule is derived. A more general approach based on the Boltzmann transport equation is also discussed. We review the conductivity of metals and include a useful collection of experimental data. The conductivity of nonuniform materials such as alloys, polycrystalline materials, composites and thin films is discussed in the context of Nordheim's rule for alloys, effective medium theories for inhomogeneous materials, and theories of scattering for thin films. We also discuss some interesting aspects of conduction in the presence of a magnetic field (the Hall effect). We present a simplified analysis of charge transport in semiconductors in a high electric field, including a modern avalanche theory (the theory of "lucky" drift). The properties of low-dimensional systems are briefly reviewed, including the quantum Hall effect.

2.1	Fundamentals: Drift Velocity, Mobility and Conductivity	20
2.2	Matthiessen's Rule	22
2.3	Resistivity of Metals	23
2.3.1	General Characteristics	23
2.3.2	Fermi Electrons	25
2.4	Solid Solutions and Nordheim's Rule	26
2.5	Carrier Scattering in Semiconductors	28
2.6	The Boltzmann Transport Equation	29
2.7	Resistivity of Thin Polycrystalline Films ..	30
2.8	Inhomogeneous Media. Effective Media Approximation	32
2.9	The Hall Effect	35
2.10	High Electric Field Transport	37
2.11	Avalanche	38
2.12	Two-Dimensional Electron Gas	39
2.13	One Dimensional Conductance	41
2.14	The Quantum Hall Effect	42
	References	44

A good understanding of charge carrier transport and electrical conduction is essential for selecting or developing electronic materials for device applications. Of particular importance are the drift mobility of charge carriers in semiconductors and the conductivity of conductors and insulators. Carrier transport is a broad field that encompasses both traditional 'bulk' processes and, increasingly, transport in low dimensional or quantized structures. In other chapters of this handbook, Baranovskii describes hopping transport in low mobility solids such as insulators, Morigaki deals with the electrical properties of amorphous semiconductors and Gould discusses in detail conduction in thin films. In this chap-

ter, we outline a semi-quantitative theory of charge transport suitable for a wide range of solids of use to materials researchers and engineers. We introduce theories of "bulk" transport followed by processes pertinent to ultra-fast transport and quantized transport in lower dimensional systems. The latter covers such phenomena as the Quantum Hall Effect, and Quantized Conductance and Ballistic Transport in Quantum Wires that has potential use in new kinds of devices. There are many more rigorous treatments of charge transport; those by *Rossiter* [2.1] and *Dugdale* [2.2] on metals, and *Nag* [2.3] and *Blatt* [2.4] on semiconductors are highly recommended.

2.1 Fundamentals: Drift Velocity, Mobility and Conductivity

Basic to the theory of the electronic structure of solids are the solutions to the quantum mechanical problem of an electron in a periodic potential known as Bloch waves. These wavefunctions are traveling waves and provide the physical basis for conduction. In the semi-classical approach to conduction in materials, an electron wavepacket made up of a superposition of Bloch waves can in principle travel unheaded in an ideal crystal. No crystal is ideal, however, and the imperfections cause scattering of the wavepacket. Since the interaction of the electron with the potential of the ions is incorporated in the Bloch waves, one can concentrate on the relatively rare scattering events which greatly simplifies the theory. The motion of the electrons between scattering events is essentially free (with certain provisos such as no interband transitions) subject only to external forces, usually applied electric or magnetic fields. A theory can then be developed that relates macroscopic and measurable quantities such as conductivity or mobility to the microscopic scattering processes. Principle in such a theory is the concept of *mean free time* τ which is the average time between scattering events. τ is also known as the *conductivity relaxation time* because it represents the time scale for the momentum gained from an external field to relax. Equivalently, $1/\tau$ is the average probability per unit time that an electron is scattered.

There are two important velocity quantities that must be distinguished. The first is the *mean speed* u or *thermal velocity* v_{th} which as the name implies is the average speed of the electrons. u is quite large being on the order of $\sqrt{3k_B T/m_e^*} \approx 10^5$ m/s for electrons in a nondegenerate semiconductor and $\sqrt{2E_F/m_e^*} \approx 10^6$ m/s for electrons in a metal, where k_B is Boltzmann's constant, T is the temperature, E_F is the Fermi energy of the metal, and m_e^* is the electron effective mass. The distance an electron travels between scattering events is called the free path. It is straightforward to show that the average or *mean free path* for an electron is simply $\ell = u\tau$. The second velocity is the mean or *drift velocity* \mathbf{v}_d (variables in boldface are vectors) which is simply the vector average over the velocities of all N electrons,

$$\mathbf{v}_d = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i. \quad (2.1)$$

With no external forces applied to the solid, the electron motion is random and thus the drift velocity is zero. When subject to external forces like an electric field, the electrons acquire a net drift velocity. Normally, the

magnitude of the drift velocity is much smaller than u so that the mean speed of the electron is not affected to any practical extent by the external forces. An exception is charge transport in semiconductors in high electric fields, where $|\mathbf{v}_d|$ becomes comparable to u .

The drift velocity gives rise to an electric current. If the density of electrons is n then the current density \mathbf{J}_e is

$$\mathbf{J}_e = -en\mathbf{v}_d \quad (2.2)$$

where e is the fundamental unit of electric charge. For the important case of an applied electric field \mathbf{E} , the solutions of the semi-classical equations give a drift velocity that is proportional to \mathbf{E} . The proportionality constant is the *drift mobility* μ_e

$$\mathbf{v}_d = -\mu_e \mathbf{E}. \quad (2.3)$$

The drift mobility might be a constant or it might depend on the applied field (usually only if the field is large). Ohm's Law defines the conductivity σ of a material $\mathbf{J} = \sigma \mathbf{E}$ resulting in a simple relation to the drift mobility

$$\sigma = en\mu_e. \quad (2.4)$$

Any further progress requires some physical theory of scattering. A useful model results from the simple assumption that the scattering randomizes the electron's velocity (taking into proper account the distribution of electrons and the Pauli Exclusion Principle). The equation of motion for the drift velocity then reduces to a simple form

$$\frac{d\mathbf{v}_d}{dt} = \frac{\mathbf{F}(t)}{m_e^*} - \frac{\mathbf{v}_d}{\tau}, \quad (2.5)$$

where $\mathbf{F}(t)$ is the sum of all external forces acting on the electrons. The effect of the scattering is to introduce a frictional term into what otherwise would be just Newton's Law. Solutions of (2.5) depend on $\mathbf{F}(t)$. In the simplest case of a constant applied electric field, the steady-state solution is trivial,

$$\mathbf{v}_d = \frac{-e\mathbf{E}\tau}{m_e^*}. \quad (2.6)$$

The conductivity and drift mobility can now be related to the scattering time [2.5],

$$\mu_e = e\tau/m_e^* \quad \text{and} \quad \sigma = ne^2\tau/m_e^*. \quad (2.7)$$

More sophisticated scattering models lead to more accurate but more complicated solutions.

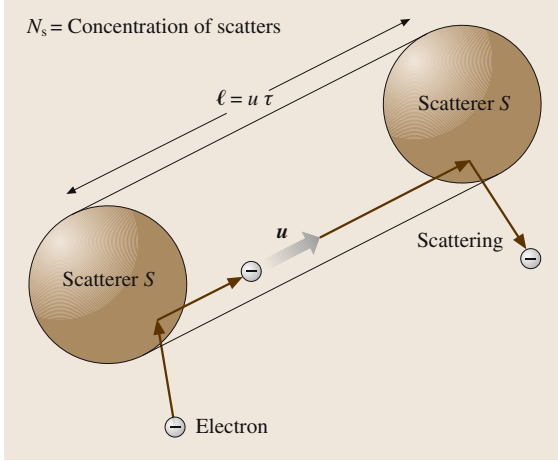


Fig. 2.1 Scattering of an electron from a scattering center. The electron travels a mean distance $\ell = u\tau$ between collisions

The simple expression (2.7) can be used to explain qualitative features of conduction in materials once a physical origin for the scattering is supplied. For any scattering site, the effective area is the *scattering cross-section* S as depicted in Fig. 2.1. The scattering cross section is related to the mean free path since the volume $S\ell$ must contain one scattering center. If there are N_S scattering centers per unit volume then

$$\ell = \frac{1}{SN_S}, \quad (2.8)$$

or, rewriting in terms of τ ,

$$\tau = \frac{1}{SN_S u}. \quad (2.9)$$

Once the cross-section for each physically relevant scattering mechanism is known then the effect on the scattering time and conductivity is readily calculated.

An overly restrictive assumption in the above analysis is that the electron's velocity is completely randomized every time it is scattered. On the other hand, suppose that ν collisions are required to completely destroy the directional velocity information. That is only after an average of ν collisions do all traces of correlation between the initial and the final velocities disappear. The *effective* mean free path ℓ_{eff} traversed by the electron until its velocity is randomized will now be larger than ℓ ; to first order $\ell_{\text{eff}} = \nu\ell$. ℓ_{eff} is termed the *effective* or the *conduction mean free path*. The corresponding

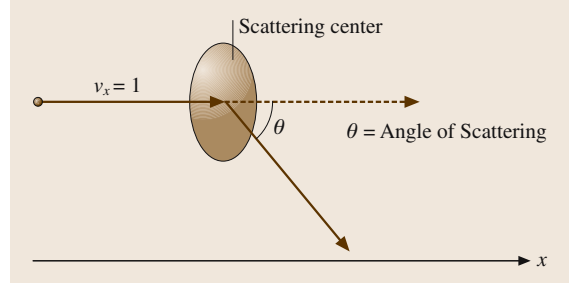


Fig. 2.2 An electron moving in the x-direction becomes scattered through an angle θ with respect to the original direction

effective scattering cross section is

$$S_{\text{eff}} = \frac{1}{N_S \ell_{\text{eff}}}. \quad (2.10)$$

The expressions for mobility and conductivity become

$$\mu = \frac{e\nu\tau}{m_e^*} = \frac{e\nu\ell}{m_e^* u} = \frac{e\ell_{\text{eff}}}{m_e^* u} = \frac{e}{m_e^* N_S S_{\text{eff}}} \quad (2.11)$$

and

$$\sigma = \frac{e^2 \nu \tau}{m_e^*} = \frac{e^2 n \nu \ell}{m_e^* u} = \frac{e^2 n \ell_{\text{eff}}}{m_e^* u} = \frac{e^2 n}{m_e^* N_S S_{\text{eff}}}. \quad (2.12)$$

Suppose that in a collision the electron is scattered at an angle θ to its original direction of travel as shown in Fig. 2.2. It is convenient to introduce a quantity $S_\theta(\theta)$, called the *differential scattering cross section*, defined so that $2\pi \sin \theta S_\theta d\theta$ represents the probability of scattering at an angle between θ and $\theta + d\theta$ with respect to the original direction. If the magnitude of the velocity is not changed then the fractional change in component of the velocity along the original direction is $1 - \cos \theta$. The average number of collisions ν required to randomize the velocity is then

$$\nu = \frac{1}{\langle 1 - \cos \theta \rangle}, \quad (2.13)$$

where the average is given by

$$\langle 1 - \cos \theta \rangle = \frac{\int_0^\pi (1 - \cos \theta) S_\theta(\theta) \sin \theta d\theta}{\int_0^\pi S_\theta(\theta) \sin \theta d\theta}. \quad (2.14)$$

The effective cross sectional area S_{eff} is then

$$S_{\text{eff}} = 2\pi \int_{\theta_m}^\pi (1 - \cos \theta) S_\theta(\theta) \sin \theta d\theta. \quad (2.15)$$

As an example, consider conduction electrons scattering from charged impurities in a nondegenerate semiconductor where small angle deviations predominate. The differential cross section for coulombic scattering from a charged impurity center with charge $+Ze$ is

$$S_{\theta}(\theta) = \frac{16k^2}{u^4} \times \frac{1}{\theta^4}; \quad k = \frac{Ze^2}{4\pi\epsilon m_e^*}, \quad (2.16)$$

where $k = Ze^2/4\pi\epsilon_0\epsilon_r m_e^*$ and ϵ_r is the relative permittivity of the semiconductor. Let r_m be the maximum effective radius of action for the impurity at which the minimum scattering angle occurs θ_m . Then the integral in (2.15) evaluates to

$$S_{\text{eff}} = \frac{2\pi k^2}{u^4} \ln \left(1 + \frac{r_m^2 u^4}{k^2} \right). \quad (2.17)$$

In a nondegenerate semiconductor, the equipartition theorem links velocity to temperature, $m_e^* u^2/2 = 3kT/2$, so

that $u \propto T^{1/2}$. Thus,

$$S_{\text{eff}} = \frac{A}{T^2} \ln(1 + BT^2), \quad (2.18)$$

where A and B are constants. The drift mobility due to scattering from ionized impurities becomes

$$\begin{aligned} \mu_I &= \frac{e}{m_e^* u S_{\text{eff}} N_I} \propto \frac{1}{T^{1/2}} \frac{T^2}{A \ln(1 + BT^2)} \frac{1}{N_I} \\ &\approx \frac{CT^{3/2}}{N_I}, \end{aligned} \quad (2.19)$$

where N_I is the density of ionized impurities and C is a new constant. At low temperatures where lattice scattering is insignificant, we expect $\mu_I \propto T^{3/2}/N_I$ for nondegenerate semiconductors.

The above semiquantitative description is sufficient to understand the basic principles of conduction. A more rigorous approach involves solving the Boltzmann charge transport equation and is addressed in Sect. 2.6.

2.2 Matthiessen's Rule

In general, the conduction electron whether in a metal or in a semiconductor can be scattered by a number of mechanisms, such as lattice vibrations, impurities, lattice defects such as dislocations, grain boundaries, vacancies, surfaces, or any other deviation from a perfectly periodic lattice. All these scattering processes increase the overall resistivity of the substance by reducing the mean scattering time. The relation between the types of scattering and the total scattering time can be obtained by considering scattering from lattice vibrations and im-

purities as shown in Figure 3.1. We define two mean free times τ_L and τ_I : τ_L is the mean free time considering only scattering from lattice vibrations (phonons) and τ_I is the mean free time considering only collisions with impurities. In a small unit of time dt , the total probability of scattering (dt/τ) is simply the sum of the probability for phonon scattering (dt/τ_L) and the probability for impurity scattering (dt/τ_I), and thus

$$\frac{1}{\tau} = \frac{1}{\tau_L} + \frac{1}{\tau_I}. \quad (2.20)$$

We have assumed that neither τ_L nor τ_I is affected by the presence of the other scattering mechanism, that is each type of scattering is independent. The above expression can be generalized to include all types of independent scattering yielding

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}, \quad (2.21)$$

where τ_i is the mean scattering time considering the i th scattering process alone. Since the drift mobility is proportional to τ , (2.21) can be written in terms of the drift mobilities determined by the various scattering mechanisms. In other words,

$$\frac{1}{\mu_d} = \sum_i \frac{1}{\mu_i} \quad (2.22)$$

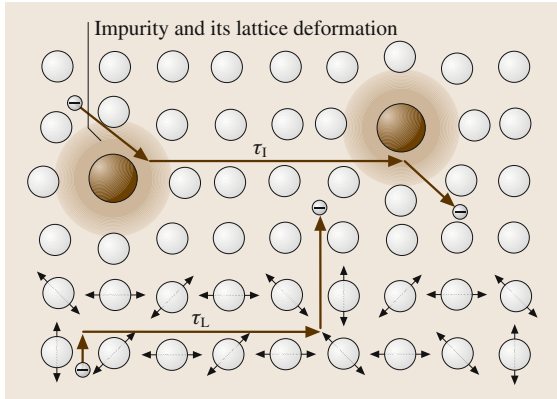


Fig. 2.3 Scattering from lattice vibrations alone with a mean scattering time τ_L , and from impurities alone with a mean scattering time τ_I

where μ_i is the drift mobility limited just by the i th scattering process. Finally, since the resistivity is inversely proportional to the drift mobility, the relation for resistivity is

$$\rho = \sum_i \rho_i \quad (2.23)$$

where ρ_i is the resistivity of the material if only the i th scattering process were active. Equation (2.23) is known as Matthiessen's rule. For nearly perfect, pure

crystals the resistivity is dominated by phonon scattering ρ_T . If impurities or defects are present, however, there are an additional resistivities ρ_I from the scattering off the impurities and ρ_D from defect scattering, and $\rho = \rho_T + \rho_I + \rho_D$.

Matthiessen's rule is indispensable for predicting the resistivities of many types of conductors. In some cases like thin films, the rule is obeyed only approximately, but it is nonetheless still useful for an initial (often quite good) estimate.

2.3 Resistivity of Metals

2.3.1 General Characteristics

The effective resistivity of a metal, by virtue of Matthiessen's rule, is normally written as

$$\rho = \rho_T + \rho_R, \quad (2.24)$$

where ρ_R is called the *residual resistivity* and is due to the scattering of electrons by impurities, dislocations, interstitial atoms, vacancies, grain boundaries and so on. The residual resistivity shows very little temperature dependence whereas ρ_T is nearly linear in absolute temperature. ρ_T will be the main resistivity term for many good-quality, pure, crystalline metals. In classical terms, we can take the thermal vibrations of a lattice atom with mass M as having a mean kinetic energy KE of $(1/2)Ma^2\omega^2$, where a and ω are the amplitude and frequency of the vibrations. This mean KE must be of the order of kT so that the amplitude $a \propto T^{1/2}$. Thus the electron scattering cross section $S = \pi a^2 \propto T$. Since the mean speed of conduction electron in a metal is the Fermi speed and is temperature independent, $\mu \propto \tau \propto 1/S \propto T^{-1}$, and hence the resistivity $\rho \propto T$. Most nonmagnetic pure metals obey this relationship except at very low temperatures. Figure 2.4 shows the resistivity of Cu as a function of temperature where above ≈ 100 K, $\rho \propto T$.

Frequently, the resistivity vs. temperature behavior of pure metals can be empirically represented by a power law of the form,

$$\rho = \rho_0 \left(\frac{T}{T_0} \right)^n, \quad (2.25)$$

where ρ_0 is the resistivity at the reference temperature, T_0 , and n is a characteristic index that best fits the data. For the nonmagnetic metals, n is close to unity whereas it is close to 2 for the magnetic metals Fe and Ni [2.5]. Figure 2.5 shows ρ vs. T for various

metals. Table 2.1 summarizes the values ρ_0 and n for various metals.

As apparent from Fig. 2.4, below ≈ 100 K, the $\rho \propto T$ behavior fails, and $\rho \propto T^5$. The reason is that, as the temperature is lowered, the scattering by phonons becomes less efficient, and it takes many more collisions to fully randomize the initial velocity of the electron. The mean number of collisions ν required to randomize the velocity scales with T^{-2} [2.5], and at low temperatures,

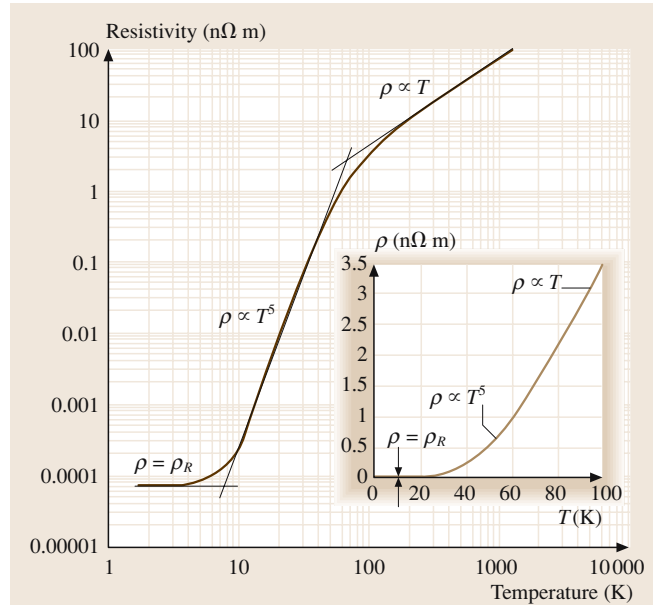


Fig. 2.4 The resistivity of copper from low to high temperatures (near its melting temperature, 1358 K) on a log–log plot. Above about 100 K, $\rho \propto T$, whereas at low temperatures, $\rho \propto T^5$, and at the lowest temperatures ρ approaches the residual resistivity ρ_R . The inset shows the ρ vs. T behavior below 100 K on a linear plot. (ρ_R is too small to see on this scale.) After [2.5]

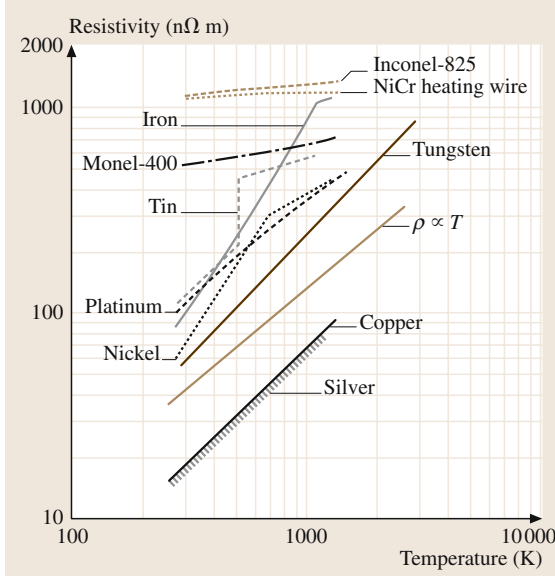


Fig. 2.5 The resistivities of various metals as a function of temperature above 0°C. Tin melts at 505 K, whereas nickel and iron go through a magnetic-to-nonmagnetic (Curie) transformation at about 627 K and 1043 K, respectively. The theoretical behavior ($\rho \propto T$) is shown for reference. After [2.5]

the concentration n_{phonon} of phonons increases as T^3 . Thus,

$$\sigma \propto v\tau \propto \frac{v}{n_{\text{phonon}}} \propto \frac{T^{-2}}{T^3} \propto T^{-5} \quad (2.26)$$

which explains the low-temperature $\rho - T$ behavior in Fig. 2.4. The low temperature $\rho \propto T^5$ and high temperature $\rho \propto T$ regimes are roughly separated by the *Debye temperature* T_D . For $T > T_D$ we expect $\rho \propto T$, and for $T < T_D$ we expect $\rho \propto T^5$.

In the case of metals with impurities and for alloys, we need to include the ρ_R contribution to the overall resistivity. For $T > T_D$, the overall resistivity is

$$\rho \approx AT + \rho_R, \quad (2.27)$$

where A is a constant, and the AT term in (2.27) arises from scattering from lattice vibrations. Normally, ρ_R has very little temperature dependence, and hence very roughly ρ vs. T curves shift to higher values as ρ_R is increased due to the addition of impurities, alloying or cold working the sample (mechanical deformation that generates dislocations) as illustrated for Cu–Ni alloys in Fig. 2.6.

Resistivity vs. temperature behavior of nearly all metals is characterized by the *temperature coefficient of resistivity* (TCR) α_0 which is defined as the fractional change in the resistivity per unit temperature increase at the reference temperature T_0 , i. e.

$$\alpha_0 = \frac{1}{\rho_0} \left(\frac{d\rho}{dT} \right)_{T=T_0}, \quad (2.28)$$

where ρ_0 is the resistivity at the reference temperature T_0 , usually at 273 K (0°C) or 293 K (20°C), $d\rho = \rho - \rho_0$ is the change in the resistivity due to a small increase, $dT = T - T_0$, in temperature. Assuming that α_0 is temperature independent over a small range from T_0 to T , we can integrate (2.28), which leads to the well known equation,

$$\rho = \rho_0[1 + \alpha_0(T - T_0)]. \quad (2.29)$$

Equation (2.29) is actually only valid when α_0 is *constant* over the temperature range of interest which requires (2.27) to hold. Over a limited temperature range this will usually be the case. Although it is not obvious from (2.28), we should, nonetheless, note that α_0 depends on the reference temperature, T_0 by virtue of ρ_0 depending on T_0 .

It is instructive to mention that if $\rho \approx AT$ as we expect for an ideal pure metal, then $\alpha_0 = T_0^{-1}$. If we take the reference temperature T_0 as 273 K (0°C), then α_0 should ideally be $1/(273 \text{ K})$ or $3.66 \times 10^{-3} \text{ K}^{-1}$. Examination of α_0 for various metals shows that $\rho \propto T$ is not a bad approximation for some of the familiar

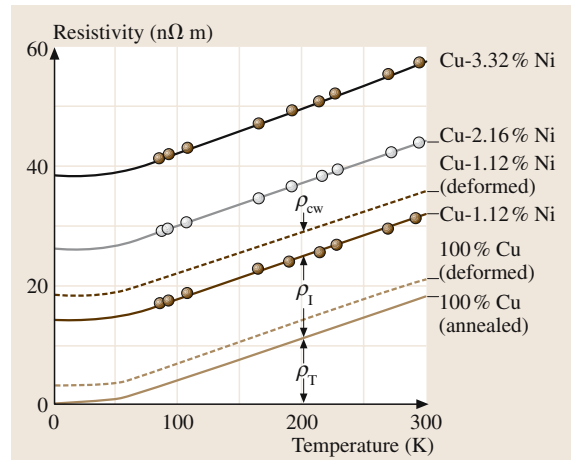


Fig. 2.6 Resistivities of annealed and cold-worked (deformed) copper containing various amounts of Ni (given in atomic percentages) versus temperature

Table 2.1 Resistivities at 293 K (20 °C) ρ_0 and thermal coefficients of resistivity α_0 at 0–100 °C for various metals. The resistivity index n in $\rho = \rho_0(T/T_0)^n$ is also shown. Data was compiled from [2.6, 7]

Metal	$\rho_0(n\Omega\text{m})$	n	$\alpha_0 \times 10^{-3} (\text{K}^{-1})$
Aluminium, Al	26.7	1.20	4.5
Barium, Ba	600	1.57	
Beryllium, Be	33	1.84	9
Bismuth, Bi	1170	0.98	4.6
Cadmium, Cd	73	1.16	4.3
Calcium, Ca	37	1.09	4.57
Cerium, Ce	854	1.35	8.7
Cesium, Cs	200	1.16	4.8
Cromium, Cr	132	1.04	2.14
Cobalt, Co	63	1.80	6.6
Copper, Cu	16.94	1.15	4.3
Gallium, Ga	140		
Gold, Au	22	1.11	4
Hafnium, Ha	322	1.20	4.4
Indium, In	88	1.40	5.2
Iridium, Ir	51	1.17	4.5
Iron, Fe	101	1.73	6.5
Lead, Pb	206	1.13	4.2
Lithium, Li	92.9	1.23	4.35
Magnesium, Mg	4.2	1.09	4.25
Molybdenum, Mo	57	1.26	4.35
Nickel, Ni	69	1.64	6.8
Niobium, Nb	160	0.80	2.6
Osmium, Os	88	1.10	4.1
Palladium, Pd	108	0.96	4.2
Platinum, Pt	105.8	1.02	3.92
Potassium, K	68	1.38	5.7
Rhodium, Rh	47	1.21	4.4
Rubidium, Rb	121	1.41	4.8
Ruthenium, Ru	77	1.15	4.1
Silver, Ag	16.3	1.13	4.1
Sodium, Na	47	1.31	5.5
Strontium, Sr	140	0.99	3.2
Tantalum, Ta	135	1.01	3.5
Tin, Sn	126	1.4	4.6
Titanium, Ti	540	1.27	3.8
Tungsten, W	54	1.26	4.8
Vanadium, V	196	1.02	3.9
Zinc, Zn	59.6	1.14	4.2
Zirconium, Zr	440	1.03	4.4

pure metals used as conductors, e.g. Cu, Al, Au, but fails badly for others, such as indium, antimony

and, in particular, the magnetic metals, e.g. iron and nickel.

Frequently we are given α_0 at a temperature T_0 , and we wish to use some other reference temperature, say T_0' , that is, we wish to use ρ_0' and α_0' for ρ_0 and α_0 respectively in (2.29) by changing the reference from T_0 to T_0' . Then we can find α_1 from α_0 ,

$$\alpha_0' = \frac{\alpha_0}{1 + \alpha_0(T_0' - T_0)}$$

and $\rho = \rho_0' [1 + \alpha_0'(T - T_0')] . \quad (2.30)$

For example, for Cu $\alpha_0 = 4.31 \times 10^{-3} \text{ K}^{-1}$ at $T_0 = 0^\circ\text{C}$, but it is $\alpha_0 = 3.96 \times 10^{-3} \text{ K}^{-1}$ at $T_0 = 20^\circ\text{C}$. Table 2.1 summarizes α_0 for various metals.

2.3.2 Fermi Electrons

The electrical properties of metals depend on the behavior of the electrons at the Fermi surface. The electron states at energies more than a few kT below E_F are almost fully occupied. The Pauli exclusion principle requires that an electron can only be scattered into an empty state, and thus scattering of deep electrons is highly suppressed by the scarcity of empty states (scattering where the energy changes by more than a few kT is unlikely). Only the electrons near E_F undergo scattering. Likewise, under the action of an external field, only the electron occupation near E_F is altered. As a result, the density of states (DOS) near the Fermi level is most important for the metal electrical properties, and only those electrons in a small range ΔE around E_F actually contribute to electrical conduction. The density of these electrons is approximately $g(E_F)\Delta E$ where $g(E_F)$ is the DOS at the Fermi energy. From simple arguments, the overall conductivity can be shown to be [2.5]

$$\sigma = \frac{1}{3} e^2 v_F^2 \tau g(E_F) , \quad (2.31)$$

where v_F is the Fermi speed and τ is the scattering time of these Fermi electrons. According to (2.31), what is important is the density of states at the Fermi energy, $g(E_F)$. For example, Cu and Mg are metals with valencies I and II. Classically, Cu and Mg atoms each contribute 1 and 2 conduction electrons respectively into the crystal. Thus, we would expect Mg to have higher conductivity. However, the Fermi level in Mg is where the top tail of the 3p-band overlaps the bottom tail of the 3s band where the density of states is small. In Cu, on the other hand, E_F is nearly in the middle of the 4s band where the density of states is high. Thus, Mg has a lower conductivity than Cu.

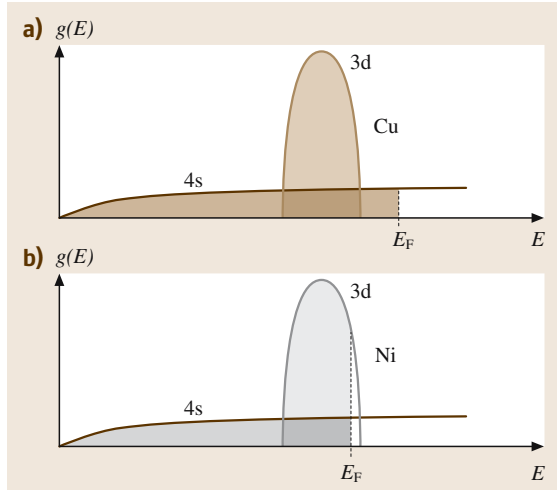


Fig. 2.7a,b Simplified energy band diagrams around E_F for copper (a) and nickel (b)

The scattering time τ in (2.31) assumes that the scattered electrons at E_F remain in the same energy band as,

for example, in Cu, whose simplified energy band diagram around E_F is shown in Fig. 2.7a. In certain metals, there are two different energy bands that overlap at E_F . For example, in Ni, 3d and 4s bands overlap at E_F as shown in Fig. 2.7b. An electron can be scattered from the 4s to the 3d band and vice versa. Electrons in the 3d band have very low drift mobilities and effectively do not contribute to conduction so that only $g(E_F)$ of the 4s band operates in (2.31). Since 4s to 3d band is an additional scattering mechanism, by virtue of Matthiessen's rule, the effective scattering time τ for the 4s band electrons is shortened and hence σ from (2.31) is smaller. Thus, Ni has poorer conductivity than Cu.

Equation (2.31) does not assume a particular density of states model. If we now apply the *free electron model* for $g(E_F)$, and also relate E_F to the total number of conduction electrons per unit volume n , we would find that the conductivity is the same as that in the Drude model, that is

$$\sigma = \frac{e^2 n \tau}{m_e} . \quad (2.32)$$

2.4 Solid Solutions and Nordheim's Rule

In an *isomorphous alloy* of two metals, i.e. a binary alloy which forms a binary *substitutional solid solution*, an additional mechanism of scattering appears, the *scattering off solute phase atoms*. This scattering contributes to lattice scattering, and therefore increases the overall resistivity. An important semi-empirical equation which can be used to predict the resistivity of an alloy is Nordheim's rule. It relates the impurity part of the resistivity ρ_I to the atomic fraction X of solute atoms in a solid solution via

$$\rho_I = CX(1 - X) , \quad (2.33)$$

where the constant C is termed the Nordheim coefficient and represents the effectiveness of the solute atom in increasing the resistivity. Nordheim's rule was originally derived for crystals. Combining Nordheim's rule with Matthiessen's rule (2.23), the resistivity of an alloy of composition X should follow

$$\rho = \rho_{\text{matrix}} + CX(1 - X) , \quad (2.34)$$

where ρ_{matrix} is the resistivity of the matrix due to scattering from thermal vibrations and from other defects, in the absence of alloying elements.

Nordheim's rule assumes that the solid solution has the solute atoms randomly distributed in the lattice. For sufficiently small amounts of impurity, experiments show that the increase in the resistivity ρ_I is nearly always simply proportional to the impurity concentration X , that is, $\rho_I \propto X$. For dilute solutions, Nordheim's rule predicts the same linear behavior, that is, $\rho_I = CX$ for $X \ll 1$.

Originally the theoretical model for ρ_I was developed by Nordheim [2.8] by assuming that the solute atoms simply perturb the periodic potential and thereby increase the probability of scattering. Quantum mechanical calculations for electron scattering within a single band, such as the *s*-band, at E_F show that

$$\rho_I \propto g(E_F) V_{\text{scatter}}^2 X(1 - X) , \quad (2.35)$$

where $g(E_F)$ is the DOS at E_F , and V_{scatter} is matrix element for scattering from one wavefunction to another at the Fermi surface in the same band, which for an *s*-band is

$$V_{\text{scatter}} = \langle \psi_s^* | \Delta V | \psi_s \rangle , \quad (2.36)$$

Table 2.2 Nordheim coefficients (at 20 °C) for dilute alloys obtained from $\rho_i = CX$ and $X < 1$ at.%. Note: For many isomorphous alloys, C may be different at higher concentrations; that is, it may depend on the composition of the alloy [2.7, 9]. Maximum solubility data from [2.10]

Solute in solvent (element in matrix)	Nordheim coefficient (nΩ m)	Maximum solubility at 25 °C (at.%)
Au in Cu matrix	5500	100
Mn in Cu matrix	2900	24
Ni in Cu matrix	1250	100
Sn in Cu matrix	2900	0.6
Zn in Cu matrix	300	30
Cu in Au matrix	450	100
Mn in Au matrix	2410	25
Ni in Au matrix	790	100
Sn in Au matrix	3360	5
Zn in Au matrix	950	15

where ΔV is the difference between the potentials associated with solvent and solute atoms, and ψ_s is the wavefunction of an electron in the s-band at E_F . It is clear that C is only independent of X if $g(E_F)$ and V_{scatter} remain the same for various X which may not be true. For example, if the effective number of free electrons increases with X , E_F will be shifted higher, and C will not be constant.

Table 2.2 lists some typical Nordheim coefficients for various additions to copper and gold. The value of the Nordheim coefficient depends on the type of solute and the solvent. A solute atom that is drastically different in size to the solvent atom will result in a bigger increase in ρ_l and will therefore lead to a larger C . An important assumption in Nordheim's rule in (2.33) is that the alloying does not significantly vary the number of conduction electrons per atom in the alloy. Although this will be true for alloys with the same valency, that is, from the same column in the Periodic Table (e.g., Cu–Au, Ag–Au), it will not be true for alloys of different valency, such as Cu and Zn. In pure copper, there is just one conduction electron per atom, whereas each Zn atom can donate two conduction electrons. As the Zn content in brass is increased, more conduction electrons become available per atom. Consequently, the resistivity predicted by (2.34) at high Zn contents is greater than the actual value because C refers to dilute alloys. To get the correct resistivity from (2.34) we have to lower C , which is equivalent to using an *effective* Nordheim coefficient

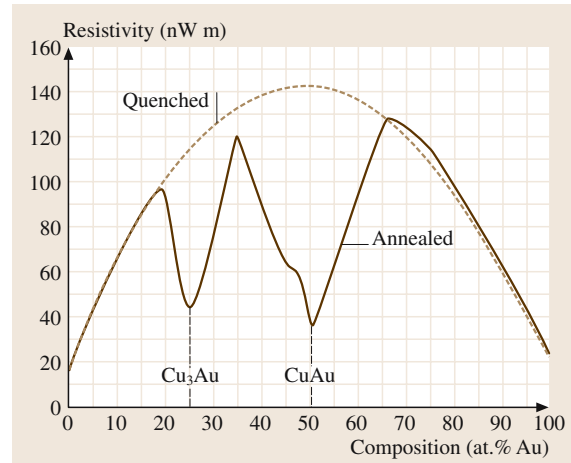


Fig. 2.8 Electrical resistivity versus composition at room temperature in Cu–Au alloys. The quenched sample (dashed curve) is obtained by quenching the liquid, and it has the Cu and Au atoms randomly mixed. The resistivity obeys the Nordheim rule. On the other hand, when the quenched sample is annealed or the liquid slowly cooled (solid curve), certain compositions (Cu₃Au and CuAu) result in an ordered crystalline structure in which Cu and Au atoms are positioned in an ordered fashion in the crystal and the scattering effect is reduced

C_{eff} that decreases as the Zn content increases. In other cases, for example, in Cu–Ni alloys, we have to increase C at high Ni concentrations to account for additional electron scattering mechanisms that develop with Ni addition. Nonetheless, the Nordheim rule is still useful for predicting the resistivities of dilute alloys, particularly in the low-concentration region.

In some solid solutions, at some concentrations of certain binary alloys, such as 75% Cu–25% Au and 50% Cu–50% Au, the annealed solid has an orderly structure; that is, the Cu and Au atoms are not randomly mixed, but occupy regular sites. In fact, these compositions can be viewed as a pure compound – like the solids Cu₃Au and CuAu. The resistivities of Cu₃Au and CuAu will therefore be less than the same composition random alloy that has been quenched from the melt. As a consequence, the resistivity ρ versus composition X curve does not follow the dashed parabolic curve throughout; rather, it exhibits sharp falls at these special compositions, as illustrated Fig. 2.8. The *effective media approximation* may be used as an effective tool to estimate the resistivities of inhomogeneous media.

2.5 Carrier Scattering in Semiconductors

At low electric fields, ionized impurity scattering and phonon scattering predominate. Other types of scattering include carrier-carrier scattering, inter-valley scattering, and neutral impurity scattering; these may generally be ignored to a first approximation.

For phonon scattering, both *polar* and *non-polar phonon scattering* should be considered. In polar scattering, short wavelength oscillations of atoms on different sub-lattices vibrating out of phase produce an effective dipole moment proportional to the bond polarity. Since such vibrational modes are optically active (since this dipole moment can interact with an incident electromagnetic field), this type of lattice scattering is usually referred to as *polar optical phonon scattering*. Since a sub-lattice is necessary for optical modes, this scattering mechanism is not present in elemental semiconductors such as Si, Ge, or diamond.

Non-polar phonon scattering comes from long wavelength oscillations in the crystal, involving small displacements of tens to thousands of atoms. The wavelength depends on the material and its elastic properties. Such modes are very similar to sound vibrations and are thus referred to as acoustic modes. The associated atomic displacements correspond to an effective built-in strain, with local change in the lattice potential, causing carrier scattering known as *deformation potential acoustic phonon scattering*. Since the change in potential is relatively small, the scattering efficiency is relatively low as compared with polar optical phonon scattering.

Each scattering process contributes to the drift mobility according to *Matthiessen's rule*

$$\frac{1}{\mu_e} = \sum_i \frac{1}{\mu_i} \quad (2.37)$$

as discussed in Sect. 2.3 above. Figure 2.9 shows the contributions of each scattering process for n-type ZnSe – a material used in optoelectronic devices. At room temperature, polar optical phonon scattering and ionized impurity scattering dominate. These processes depend on the carrier concentration. The curve for ionized impurity scattering decreases markedly with increasing carrier concentration owing to the increasing concentration of ionized donors that supply these carriers. The ionized impurity scattering mobility is roughly inversely proportional to the concentration of ionized impurities. However, as the carrier concentration increases, as in a degenerate semiconductor, the average energy per carrier also increases (i. e., carriers move faster on average) and thus carriers are less susceptible to being scattered from the

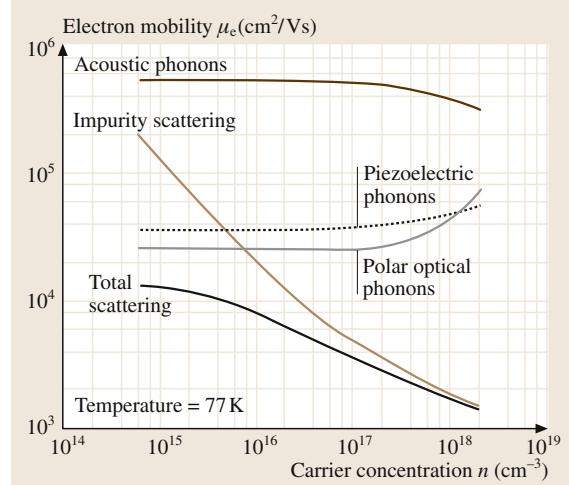


Fig. 2.9 Dependence of electron mobility on carrier concentration for ZnSe at 77 K [2.11]

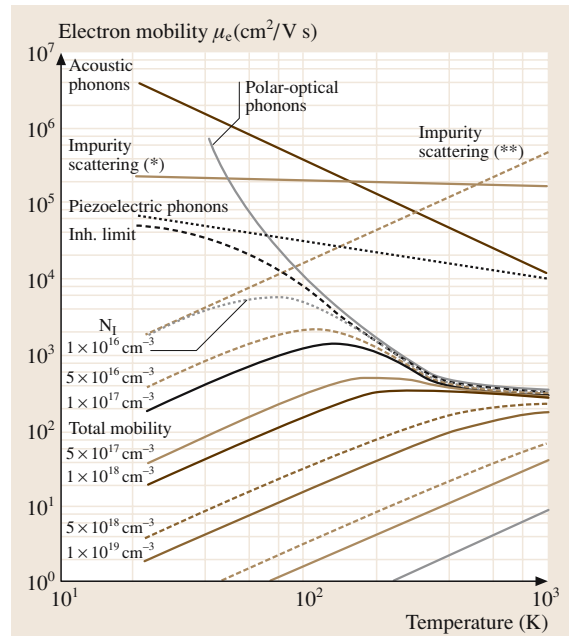


Fig. 2.10 Dependence of electron mobility on temperature and doping for ZnSe [2.11]

ionized impurity centers. In contrast, the polar phonon scattering rate is determined by the number of participating phonons which depends on the thermal energy available to create a given quantized vibrational mode.

The temperature dependence of the mobility may be estimated by considering the effect of temperature on ionized impurity and phonon scattering and combining these mechanisms using Matthiessen's rule. Phonon scattering increases strongly with increasing temperature T due to the increase in the number of phonons resulting in a $T^{-3/2}$ dependence for the polar phonon mobility. For ionized impurity scattering, increasing the temperature increases the average carrier velocity and hence increases the carrier mobility for a set concentration of ionized impurities. Once a temperature is reached

such that impurity ionization is complete, the ionized impurity based carrier mobility can be shown to increase with temperature T as approximately, $T^{+3/2}$. At low temperatures, the mobility is basically determined by ionized impurity scattering and at high temperatures by phonon scattering leading to a peaked curve. Invoking the previous discussions for the dependence of the total mobility on carrier concentration, it is clear that the peak mobility will depend on the doping level, and the peak location will shift to higher temperatures with increased doping as shown in Fig. 2.10.

2.6 The Boltzmann Transport Equation

A more rigorous treatment of charge transport requires a discussion of the *Boltzmann Transport Equation*. The electronic system is described by a distribution function $f(\mathbf{k}, \mathbf{r}, t)$ defined in such a way that the number of electrons in a six-dimensional volume element $d^3\mathbf{k}d^3\mathbf{r}$ at time t is given by $\frac{1}{4}\pi^{-3}f(\mathbf{k}, \mathbf{r}, t)d^3\mathbf{k}d^3\mathbf{r}$. In equilibrium, $f(\mathbf{k}, \mathbf{r}, t)$ depends only on energy and reduces to the Fermi distribution f_0 where the probability of occupation of states with momenta $+\mathbf{k}$ equals that for states with $-\mathbf{k}$, and $f_0(\mathbf{k})$ is symmetrical about $\mathbf{k} = 0$, giving no net charge transport. If an external field acts on the system (i.e., non-equilibrium), the occupation function $f(\mathbf{k})$ will become asymmetric in \mathbf{k} -space. If this non-equilibrium distribution function $f(\mathbf{k})$ is completely specified and appropriate boundary conditions supplied, the electronic transport properties can be completely determined by solving the *steady state Boltzmann transport equation* [2.12]

$$\mathbf{v} \cdot \nabla_{\mathbf{r}} f + \hbar \mathbf{k} \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t} \right)_c \quad (2.38)$$

where,

1. $\mathbf{v} \cdot \nabla_{\mathbf{r}} f$ represents diffusion through a volume element $d^3\mathbf{r}$ about the point \mathbf{r} in phase space due to a gradient $\nabla_{\mathbf{r}} f$,
2. $\hbar \mathbf{k} \cdot \nabla_{\mathbf{k}} f$ represents drift through a volume element $d^3\mathbf{k}$ about the point \mathbf{k} in phase space due to a gradient $\nabla_{\mathbf{k}} f$ (for example, $\hbar \dot{\mathbf{k}} = e(\mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{B})$ in the presence of electric and magnetic fields)
3. $(\partial f / \partial t)_c$ is the collision term and accounts for the scattering of electrons from a point \mathbf{k} (for example, this may be due to lattice or ionized impurity scattering).

Equation (2.38) may be simplified by using the *relaxation time approximation*

$$\left(\frac{\partial f}{\partial t} \right)_c = \frac{\Delta f}{\tau} = -\frac{f - f_0}{\tau} \quad (2.39)$$

which is based on the assumption that for small changes in f carriers return to equilibrium in a characteristic time τ , dependent on the dominant scattering mechanisms. Further simplifications of (2.38) using (2.39) applicable for low electric fields lead to a simple equation connecting the mobility μ to the *average scattering time* $\langle \tau \rangle$

$$\mu \cong \frac{e\langle \tau \rangle}{m^*} \quad (2.40)$$

The details of calculations may be found in various advanced textbooks, for example *Bube* [2.13], *Blatt* [2.4]. The average scattering time may be calculated assuming a *Maxwell-Boltzmann* distribution function and a parabolic band

$$\langle \tau \rangle = \frac{2}{3k_B T} \frac{\int_0^\infty \tau(E) E^{3/2} e^{-E/k_B T} dE}{\int_0^\infty E^{1/2} e^{-E/k_B T} dE} \quad (2.41)$$

Quantum mechanical perturbation theory can be used to calculate the carrier scattering rate for different processes i , giving,

$$\tau_i(E) = aE^{-\alpha} \quad (2.42)$$

where a and α are constants and E is the electron energy. Substituting (2.42) into (2.41) gives

$$\langle \tau_i \rangle = \frac{4a\Gamma(5/2 - \alpha)}{3\pi^{1/2}(k_B T)^\alpha} \quad (2.43)$$

in terms of the *gamma function* Γ . Using this approach, the expressions for the mobility for the case

of lattice and impurity scattering may be easily found

$$\mu_L \propto \frac{4e}{m^* \sqrt{9\pi k_B}} T^{-\frac{3}{2}}, \quad (2.44)$$

$$\mu_I \propto \frac{8ek_B^{3/2} N_I}{m^* \sqrt{\pi}} T^{+\frac{3}{2}}, \quad (2.45)$$

where N_I is the concentration of ionized impurities.

2.7 Resistivity of Thin Polycrystalline Films

Two new dominant scattering mechanisms must be considered in polycrystalline thin films – *scattering by grain boundaries* and *scattering at the surface*. The scattering by grain boundaries is schematically shown in Fig. 2.11. As a first approximation, the conduction electron may be considered free within a grain, but becomes scattered at the grain boundary. Its mean free path ℓ_{grains} is therefore roughly equal to the average grain size d . If $\lambda = \ell_{\text{crystal}}$ is the mean free path of the conduction electrons in the single crystal, then according to Matthiessen's rule

$$\frac{1}{\ell} = \frac{1}{\ell_{\text{crystal}}} + \frac{1}{\ell_{\text{grains}}} = \frac{1}{\lambda} + \frac{1}{d}. \quad (2.46)$$

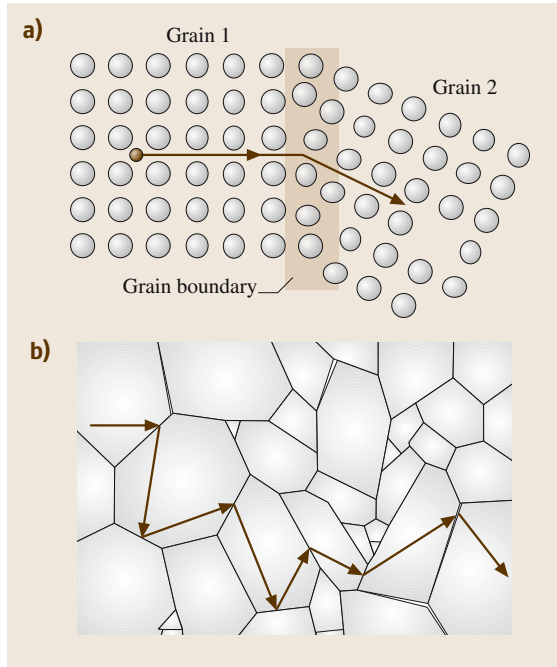


Fig. 2.11 (a) Grain boundaries cause electron scattering and therefore add to the resistivity according to Matthiessen's rule. (b) For a very grainy solid, the electron is scattered from grain boundary to grain boundary, and the mean free path is approximately equal to the mean grain diameter

The resistivity is inversely proportional to the mean free path which means the resistivity of the single crystal $\rho_{\text{crystal}} \propto 1/\lambda$ and resistivity of the polycrystalline sample $\rho \propto 1/\ell$. Thus,

$$\frac{\rho}{\rho_{\text{crystal}}} = 1 + \left(\frac{\lambda}{d} \right). \quad (2.47)$$

Figure 2.12 clearly demonstrates that even simple considerations agree well with experimental data. However, in a more rigorous theory we have to consider a number of effects. It may take more than one scattering at a grain boundary to totally randomize the velocity so that we need to calculate the effective mean free path that accounts for how many collisions are needed to randomize the velocity. There is a possibility that the electron may be totally reflected back at a grain boundary (bounce back). Let σ be the conductivity of the polycrystalline (grainy) material and σ_{crystal} be the conductivity of the bulk single crystal. Suppose that the probability of reflection at a grain boundary is R . If d is the average grain size (diameter) then the two conductivities may be related using the *Mayadas and Shatzkes*

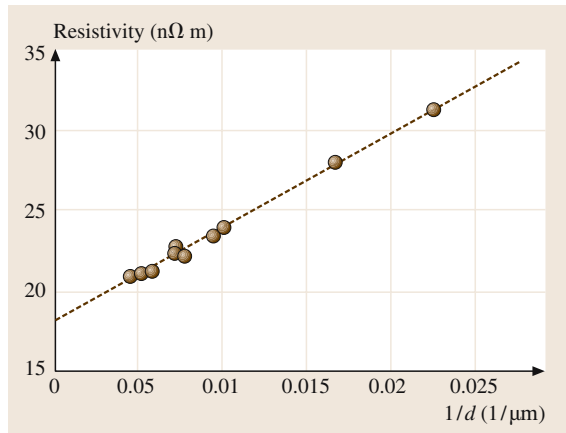


Fig. 2.12 Resistivity of Cu polycrystalline film vs. reciprocal mean grain size (diameter), $1/d$. Film thickness $D = 250\text{--}900\text{ nm}$ does not affect the resistivity. The straight line $\rho = 17.8\text{ n}\Omega\text{ m} + (595\text{ n}\Omega\text{ nm})(1/d)$ [2.14]

formula [2.15]

$$\frac{\sigma}{\sigma_{\text{crystal}}} = 1 - \frac{3}{2}\beta + 3\beta^2 - 3\beta^3 \ln\left(1 + \frac{1}{\beta}\right)$$

where $\beta = \frac{\lambda}{d} \left(\frac{R}{1-R} \right)$, (2.48)

which in $0.1 < \beta < 10$, approximates to

$$\frac{\rho}{\rho_{\text{crystal}}} \approx 1 + 1.34\beta. \quad (2.49)$$

For copper typically R values are 0.24–0.40, and somewhat a smaller R for Al. Equation (2.49) for a Cu film with $R \approx 0.3$ predicts $\rho/\rho_{\text{crystal}} \approx 1.20$ for roughly $d \approx 3\lambda$ or a grain size $d \approx 120$ nm since in the bulk crystal $\lambda \approx 40$ nm. Tellier et al. [2.17] have given extensive discussions of grain boundary scattering limited resistivity of thin films.

Scattering from the film surfaces must also be included in any resistivity calculation. It is generally assumed that the scattering from a surface is partially *inelastic*, that is the electron loses some of the velocity gained from the field. The inelastic scattering is also called *nonspecular*. (If the electron is elastically reflected from the surface just like a rubber ball bouncing from a wall, then there is no increase in the resistivity.) If the parameter p is the fraction of surface collisions which are specular (elastic) and if the thickness of film D is greater than λ , and $\sigma_{\text{bulk}} = 1/\rho_{\text{bulk}}$, then in accordance with Fuchs-Sondheimer equation [2.18, 19] the conductivity σ of the film is

$$\frac{\sigma}{\sigma_{\text{bulk}}} = 1 - \frac{3\lambda}{8D}(1-p), \quad (\lambda/D > 1). \quad (2.50)$$

If D is much shorter than ℓ ,

$$\frac{\sigma}{\sigma_{\text{bulk}}} = \frac{3D}{4\lambda} \left[\ln\left(\frac{\lambda}{D}\right) + 0.423 \right] \times (1+2p), \quad (\lambda/D \ll 1). \quad (2.51)$$

For purely nonspecular (inelastic) scattering, an approximate estimate can be obtained from

$$\frac{\rho}{\rho_{\text{bulk}}} \approx 1 + \frac{3}{8} \left(\frac{\lambda}{D} \right). \quad (2.52)$$

Figure 2.13 shows the resistivity of thin polycrystalline Cu films as a function of film thickness. From (2.50), for sufficiently small thicknesses, ρ is inversely proportional to the thickness D , which is what is observed experimentally in Fig. 2.13. The saturation at higher thicknesses is mostly due to scattering on grain boundaries.

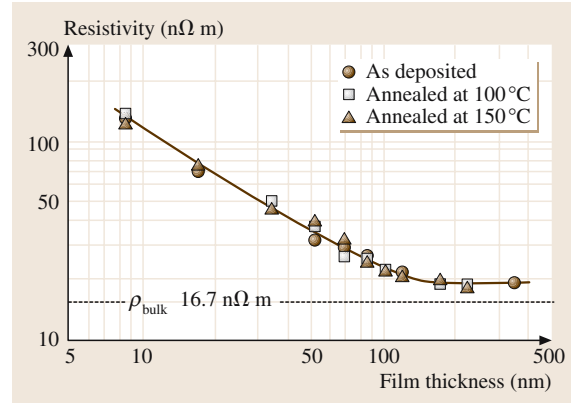


Fig. 2.13 Resistivity of Cu polycrystalline film versus film thickness D . The resistivity is mostly controlled by the surface scattering, and annealing does not significantly affect the resistivity while it reduces the crystallinity [2.16]

For elastic or specular scattering $p = 1$ and there is no change in the conductivity. The value of p depends on the film preparation method (e.g. sputtering, epitaxial growth etc.) and the substrate on which the film has been deposited. Table 2.3 summarizes the resistivities of thin Cu and Au gold films deposited by various preparation techniques. Notice the large difference between the Au films deposited on a noncrystalline glass substrate and on a crystalline mica substrate. Such differences between films are typically attributed to different values of p . The p -value can also change (increase) when the film is annealed. Obviously, the polycrystallinity of the film will also affect the resistivity as discussed above. Typically, most epitaxial thin films, unless very thin ($D \ll \ell$), deposited onto heated crystalline substrates exhibit highly specular scattering with $p = 0.9 - 1$.

It is generally very difficult to separate the effects of surface and grain boundary scattering in thin polycrystalline films; the contribution from grain boundary scattering is likely to exceed that from the surfaces. In any event, both contributions, by Matthiessen's general rule, increase the overall resistivity. Figure 2.12 shows an example in which the resistivity ρ_{film} of thin Cu polycrystalline films is due to grain boundary scattering, and thickness has no effect (D was 250 nm – 900 nm and much greater than λ). The resistivity ρ_{film} is plotted against the reciprocal mean grain size $1/d$, which then follows the expected linear behavior in (2.49). On the other hand, Fig. 2.13 shows the resistivity of Cu films as a function of film thickness D . In this case, annealing (heat treating) the films to reduce the polycrystallinity

Table 2.3 Resistivities of some thin Cu and Au films at room temperature. **PC**: Polycrystalline film; RT is room temperature; D = film thickness; d = average grain size. At RT for Cu, $\lambda = 38\text{--}40\text{ nm}$, and for Au $\lambda = 36\text{--}38\text{ nm}$. Data selectively combined from various sources, including [2.14, 16, 20–22].

Film	D (nm)	d (nm)	ρ (n Ω m)	Comment
Cu films (polycrystalline)				
Cu on TiN, W and TiW [2.14]	> 250	186	21	CVD (chemical vapor deposition). Substrate temperature 200 °C, ρ depends on d not $D = 250\text{--}900$ nm
		45	32	
Cu on 500 nm SiO ₂ [2.20]	20.5		35	Thermal evaporation, substrate at RT
	37		27	
Cu on Si (100) [2.16]	52		38	Sputtered Cu films. Annealing at 150 °C has no effect. $R \approx 0.40$ and $p \approx 0$
	100		22	
Cu on glass [2.21]	40		50	As deposited
	40		29	Annealed at 200 °C
	40		25	Annealed at 250 °C
				All thermally evaporated and PC
Au films				
Au epitaxial film on mica	30		25	Single crystal on mica. $p \approx 0.8$, specular scattering
Au PC film on mica	30		54	PC. Sputtered on mica. p is small
Au film on glass	30		70	PC. Evaporated onto glass. p is small, nonspecular scattering
Au on glass [2.22]	40		8.5	92
	40	3.8	189	

does not significantly affect the resistivity because ρ_{film} is controlled primarily by surface scattering, and is given

by (2.52). Gould in Chapt. 29 provides a more advanced treatment of conduction in thin films.

2.8 Inhomogeneous Media. Effective Media Approximation

The *effective media approximation* (**EMA**) attempts to estimate the properties of inhomogeneous mixture of two or more components using the known physical properties of each component. The general idea of any **EMA** is to substitute for the original inhomogeneous mixture some imaginary homogeneous substance – the *effective medium* (**EM**) – whose response to an external excitation is the same as that of the original mixture. The **EMA** is widely used for investigations of non-uniform objects in a variety of applications such as composite materials [2.23, 24], microcrystalline and amorphous semiconductors [2.25–28], light scattering [2.29], conductivity of dispersed ionic semiconductors [2.30] and many others.

Calculations of the conductivity and dielectric constant of two component mixtures are reviewed by Reynolds and Hough [2.31] and summarized by

Rossiter [2.1]. For such a mixture we assume that the two components α and β are randomly distributed in space with volume fractions of χ_α and $\chi_\beta = 1 - \chi_\alpha$. The dielectric properties are described by an *effective permittivity* ϵ_{eff} given by the ratio

$$\epsilon_{\text{eff}} = \langle D \rangle / \langle E \rangle , \tag{2.53}$$

where $\langle E \rangle$ is the average electric field and $\langle D \rangle$ is the average displacement field. The displacement field averaged over a large volume may be calculated from

$$\begin{aligned} \langle D \rangle &= \frac{1}{V} \int_V D dv = \frac{1}{V} \left(\int_{V_\alpha} D dv + \int_{V_\beta} D dv \right) \\ &= \chi_\alpha \langle D_\alpha \rangle + \chi_\beta \langle D_\beta \rangle , \end{aligned} \tag{2.54}$$

Table 2.4 Mixture rules for randomly oriented particles

Particle shape	Mixture rule	Factors in (2.58)		References
		A	ϵ^*	
Spheres	$\frac{\epsilon_{\text{eff}} - \epsilon_\beta}{\epsilon_{\text{eff}} + 2\epsilon_\beta} = \chi_\alpha \frac{\epsilon_\alpha - \epsilon_\beta}{\epsilon_\alpha + 2\epsilon_\beta}$	$\frac{1}{3}$	ϵ_2	[2.32–36]
Spheres	$\frac{\epsilon_{\text{eff}} - \epsilon_\beta}{3\epsilon_\beta} = \chi_\alpha \frac{\epsilon_\alpha - \epsilon_\beta}{\epsilon_\alpha + 2\epsilon_\beta}$	$\frac{1}{3}$	ϵ_2	[2.37]
Spheres	$\chi_\alpha \frac{\epsilon_\alpha - \epsilon_{\text{eff}}}{\epsilon_\alpha + 2\epsilon_{\text{eff}}} + \chi_\beta \frac{\epsilon_\beta - \epsilon_{\text{eff}}}{\epsilon_\beta + 2\epsilon_{\text{eff}}} = 0$	$\frac{1}{3}$	ϵ_{eff}	[2.38]
Spheres	$\frac{\epsilon_{\text{eff}} - \epsilon_\beta}{3\epsilon_{\text{eff}}} = \chi_\alpha \frac{\epsilon_\alpha - \epsilon_\beta}{\epsilon_\alpha + 2\epsilon_\beta}$	$\frac{1}{3}$	ϵ_{eff}	[2.39]
Spheroids	$\epsilon_{\text{eff}} = \epsilon_\beta + \frac{\chi_\alpha}{3(1 - \chi_\alpha)} \sum_{n=1}^3 \frac{\epsilon_\alpha - \epsilon_{\text{eff}}}{1 + A \left(\frac{\epsilon_\alpha}{\epsilon_{\text{eff}}} - 1 \right)}$	A	ϵ_2	[2.40]
Spheroids	$\epsilon_{\text{eff}} = \epsilon_\beta + \frac{\chi_\alpha}{3} \sum_{n=1}^3 \frac{\epsilon_\alpha - \epsilon_{\text{eff}}}{1 + A \left(\frac{\epsilon_\alpha}{\epsilon_{\text{eff}}} - 1 \right)}$	A	ϵ_{eff}	[2.41]
Lamellae	$\epsilon_{\text{eff}}^2 = \frac{2(\epsilon_\alpha \chi_\alpha - \epsilon_\beta \chi_\beta) - \epsilon_{\text{eff}}}{\frac{\epsilon_\alpha}{\chi_\alpha} + \frac{\epsilon_\beta}{\chi_\beta}}$	0	ϵ_{eff}	[2.38]
Rods	$5\epsilon_{\text{eff}}^3 + (5\epsilon'_p - 4\epsilon_p)\epsilon_{\text{eff}}^2 - (\chi_\alpha \epsilon_\alpha^2 + 4\epsilon_\alpha \epsilon_\beta + \chi_\beta \epsilon_\beta^2) - \epsilon_\alpha \epsilon_\beta \epsilon_p = 0$ where $\frac{1}{\epsilon'_p} = \frac{\chi_\alpha}{\epsilon_\beta} + \frac{\chi_\beta}{\epsilon_\alpha}$ and $\frac{1}{\epsilon_p} = \frac{\chi_\alpha}{\epsilon_\alpha} + \frac{\chi_\beta}{\epsilon_\beta}$	$\frac{1}{2}$	ϵ_{eff}	[2.42]

where $\langle D_\alpha \rangle$ and $\langle D_\beta \rangle$ are the average displacements fields inside regions of the respective components and V_α and V_β are their volumes. Likewise the electric field is given by

$$\langle E \rangle = \chi_\alpha \langle E_\alpha \rangle + \chi_\beta \langle E_\beta \rangle. \quad (2.55)$$

From (2.53) one gets

$$\epsilon_{\text{eff}} = \epsilon_\beta + (\epsilon_\alpha - \epsilon_\beta) \chi_\alpha f_\alpha \quad (2.56)$$

or

$$(\epsilon_{\text{eff}} - \epsilon_\alpha) \chi_\alpha f_\alpha + (\epsilon_{\text{eff}} - \epsilon_\beta) \chi_\beta f_\beta = 0 \quad (2.57)$$

where ϵ_α and ϵ_β are the permittivities of the components and $f_\alpha = \langle E_\alpha \rangle / \langle E \rangle$ and $f_\beta = \langle E_\beta \rangle / \langle E \rangle$ are so-called *field factors*. The choice between (2.56) and (2.57) depends on particle geometry. Equation (2.56) is better when the particles of component are dispersed in a continuous medium β . Equation (2.57) is preferred when the particle size of the two components is of the same order of magnitude.

The field factors can be calculated analytically only for phase regions with special specific geometries. The field factor for ellipsoids is given by (Stratton [2.43])

$$f_\alpha = \sum_{i=1}^3 \frac{\cos^2 \alpha_i}{1 + A_i \left(\frac{\epsilon_\alpha}{\epsilon^*} - 1 \right)} \quad (2.58)$$

where α_i are the angles between the ellipsoid axes and the applied field and A_i depends upon the axial ratios of the ellipsoids subject to the condition that

$$\sum_{i=1}^3 A_i = 1.$$

For a spheroid, $A_2 = A_3 = A$ and $A_1 = 1 - 2A$. For a random orientation of spheroids $\cos^2 \alpha_1 = \cos^2 \alpha_2 = \cos^2 \alpha_3 = \frac{1}{3}$. For the case of long particles with aligned axes $\cos^2 \alpha_1 = \cos^2 \alpha_2 = \frac{1}{2}$ and $\cos^2 \alpha_3 = 0$. The values of parameters entering (2.58) can be found in Table 2.4 which shows a set of *mixture rules*, i.e.

Table 2.5 Mixture rules for partially oriented particles

Particle shape	Formula	Factors in (2.58)				References
		A	ε^*	$\cos \alpha_1 = \cos \alpha_2$	$\cos \alpha_3$	
Parallel cylinders	$\frac{\varepsilon_{\text{eff}} - \varepsilon_\beta}{\varepsilon_{\text{eff}} + \varepsilon_\beta} = \chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_\beta}{\varepsilon_\alpha + \varepsilon_\beta}$	$\frac{1}{2}$	ε_2	$\frac{1}{2}$	0	[2.35, 36]
Parallel cylinders	$\chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_{\text{eff}}}{\varepsilon_\alpha + \varepsilon_{\text{eff}}} + \chi_\beta \frac{\varepsilon_\beta - \varepsilon_{\text{eff}}}{\varepsilon_\beta + \varepsilon_{\text{eff}}}$	$\frac{1}{2}$	ε_{eff}	$\frac{1}{2}$	0	[2.38]
Parallel lamellae (with two axes randomly oriented)	$\varepsilon_{\text{eff}}^2 = \frac{\varepsilon_\alpha \chi_\alpha + \varepsilon_\beta \chi_\beta}{\frac{\varepsilon_\alpha}{\chi_\alpha} + \frac{\varepsilon_\beta}{\chi_\beta}}$	0	ε_{eff}	$\frac{1}{2}$	0	[2.38]
Lamellae with all axes aligned (current lines are perpendicular to lamellae planes)	$\frac{1}{\varepsilon_{\text{eff}}} = \frac{\chi_\alpha}{\varepsilon_\alpha} + \frac{\chi_\beta}{\varepsilon_\beta}$	0	ε_{eff}	0	1	[2.44]
Lamellae with all axes aligned (current lines are parallel to lamellae planes)	$\varepsilon_{\text{eff}} = \varepsilon_\alpha \chi_\alpha + \varepsilon_\beta \chi_\beta$	0	ε_{eff}	1	0	[2.45, 46]
Spheroids with all axes aligned (current lines are parallel to one of the axes)	$\varepsilon_{\text{eff}} = \varepsilon_\beta + \frac{\chi_\alpha (\varepsilon_\alpha - \varepsilon_\beta)}{1 + A \left(\frac{\varepsilon_\alpha}{\varepsilon_\beta} - 1 \right)}$	A	ε_2	0	1	[2.47]
Spheroids with all axes aligned (current lines are parallel to one of the axes)	$\frac{\varepsilon_{\text{eff}}}{\varepsilon_\beta} = 1 + \frac{\chi_\alpha}{\left(\frac{\varepsilon_\alpha}{\varepsilon_\beta} - 1 \right)^{-1} + A \chi_\beta}$	A	ε_2	0	1	[2.48]

Table 2.6 Conductivity / resistivity mixture rules

Particle shape	Formula	Commentary
Lamellae with all axes aligned (current lines are perpendicular to lamellae planes)	$\rho_{\text{eff}} = \chi_\alpha \rho_\alpha + \chi_\beta \rho_\beta$	<i>Resistivity mixture rule:</i> ρ_α and ρ_β are the resistivities of two phases and ρ_{eff} is the effective resistivity of mixture
Lamellae with all axes aligned (current lines are parallel to lamellae planes)	$\sigma_{\text{eff}} = \chi_\alpha \sigma_\alpha + \chi_\beta \sigma_\beta$	<i>Conductivity mixture rule:</i> σ_α and σ_β are the conductivities of two phases and σ_{eff} is the effective conductivity of mixture
Small spheroids (α -phase) in medium (β -phase)	$\rho_{\text{eff}} = \rho_\beta \frac{(1 + \frac{1}{2} \chi_\alpha)}{(1 - \chi_\alpha)}$	$\rho_\alpha > 10 \rho_\beta$
Small spheroids (α -phase) in medium (β -phase)	$\rho_{\text{eff}} = \rho_\beta \frac{(1 - \chi_\alpha)}{(1 + 2 \chi_\alpha)}$	$\rho_\alpha < 0.1 \rho_\beta$

a set of formulae allowing one to calculate ε_{eff} for some specific cases (such as spheres, rods, lamellae, etc.). The presence of some degree of orientation somewhat simplifies the calculations as shown in the Table 2.5.

The same formulae can be used to calculate the conductivity of mixtures by substituting the appropriate conductivity σ for ε . For some special cases, the mixture rules of Table 2.5 lead to very simple formulae which allows one to calculate the conductivity of inhomogeneous alloys with those specific geometries. These *mixture rules* are summarized in Table 2.6.

The most general approach to calculating the effective dielectric permittivity comes from

$$\varepsilon_{\text{eff}} = \varepsilon_2 \left(1 - \chi_\alpha \int_0^1 \frac{G(L)}{t - L} dL \right) \quad (2.59)$$

where $t = \varepsilon_2 / (\varepsilon_2 - \varepsilon_1)$ and $G(L)$ is the *spectral function* which describes the geometry of particles. The advantage of the spectral representation is that it distinguishes between the influence of geometrical quantities and that of the dielectric properties of the components on the effective behavior of the sys-

Table 2.7 Mixture rules and corresponding spectral functions $G(L)$

Mixture rule by <i>Bruggemann</i> [2.38]: $\chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_{\text{eff}}}{\varepsilon_\alpha + 2\varepsilon_{\text{eff}}} + \chi_\beta \frac{\varepsilon_\beta - \varepsilon_{\text{eff}}}{\varepsilon_\beta + 2\varepsilon_{\text{eff}}} = 0$	
$G(L) = \frac{3\chi_\alpha - 1}{2\chi_\alpha} \delta(L) \Theta(3\chi_\alpha - 1) + \frac{3}{4\pi\chi_\alpha L} \sqrt{(L - L^-)(L^+ - L)} \Theta(L - L^-) \Theta(L^+ - L)$	
where $L^{+/-} = \frac{1}{3} \left(1 + \chi_\alpha \pm 2\sqrt{2\chi_\alpha - 2\chi_\alpha^2} \right)$	
Mixture rule by <i>Maxwell-Garnett</i> [2.49]: $\frac{\varepsilon_{\text{eff}} - \varepsilon_\beta}{\varepsilon_{\text{eff}} + 2\varepsilon_\beta} = \chi_\alpha \frac{\varepsilon_\alpha - \varepsilon_\beta}{\varepsilon_\alpha + 2\varepsilon_\beta}$	
$G(L) = \delta \left(L - \frac{1 - \chi_\alpha}{3} \right)$	
Mixture rule by <i>Looyenga</i> [2.50]: $\varepsilon_{\text{eff}}^{1/3} = \chi_\alpha \varepsilon_\alpha^{1/3} + \chi_\beta \varepsilon_\beta^{1/3}$	
$G(L) = \chi_\alpha^2 \delta(L) + \frac{3\sqrt{3}}{2\pi} \left(\chi_\beta^2 \left \frac{L-1}{L} \right ^{1/3} + \chi_\alpha \chi_\beta \left \frac{L-1}{L} \right ^{2/3} \right)$	
Mixture rule by <i>Monecke</i> [2.51]: $\varepsilon_{\text{eff}} = \frac{2(\chi_\alpha \varepsilon_\alpha + \chi_\beta \varepsilon_\beta)^2 + \varepsilon_\alpha \varepsilon_\beta}{(1 + \chi_\alpha) \varepsilon_\alpha + (2 - \chi_\alpha) \varepsilon_\beta}$	
$G(L) = \frac{2\chi_\alpha}{1 + \chi_\alpha} \delta(L) + \frac{1 - \chi_\alpha}{1 + \chi_\alpha} \delta \left(L - \frac{1 + \chi_\alpha}{3} \right)$	
Mixture rule for hollow sphere equivalent by <i>Bohren and Huffman</i> [2.52]	
$\varepsilon_{\text{eff}} = \varepsilon_\alpha \frac{(3-2f)\varepsilon_\beta + 2f\varepsilon_\alpha}{f\varepsilon_\beta + (3-f)\varepsilon_\alpha}$	
$G(L) = \frac{2}{3-f} \delta(L) + \frac{1-f}{3-f} \delta \left(L - \frac{3-f}{3} \right)$	
where $f = 1 - \frac{r_i^3}{r_o^3}$ and $r_{i/o}$ is the inner/outer radius of the sphere	

tem. Although the spectral function $G(L)$ is generally unknown for an arbitrary two-phase composite, it's analytically known or can be numerically derived

for any existing mixture rule. Examples of spectral functions and corresponding solutions are shown in Table 2.7.

2.9 The Hall Effect

The Hall effect is closely related to the phenomenon of conductivity and is observed as the occurrence of a voltage appearing across a conductor carrying an electric current in a magnetic field. The schematic of the experiment is shown in the Fig. 2.14. The effect is often characterized by the *Hall coefficient*

$$R_H = \frac{E_y}{J_x B_z}, \quad (2.60)$$

where E_y is the Hall effect electric field in the y -direction, J_x is the current density in the

x -direction and B_z is magnetic field in the z -direction.

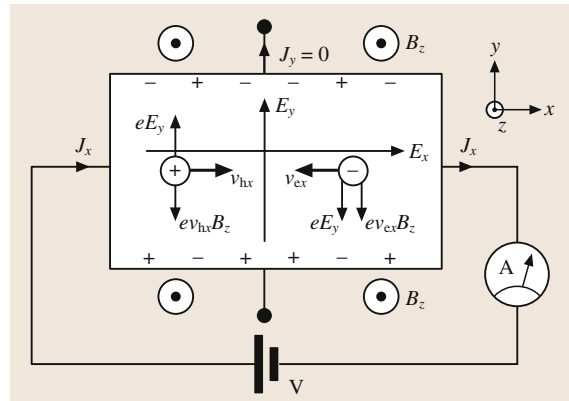


Fig. 2.14 Hall effect for ambipolar conduction. The magnetic field B_z is out of the plane of the paper. Both electrons and holes are deflected toward the bottom surface of the conductor and so the Hall voltage depends on the relative mobilities and concentrations of electrons and holes

The *Hall effect for ambipolar conduction* in a sample where there are both negative and positive charge carriers, e.g. electrons and holes in a semiconductor, involves not only the concentrations of electrons and holes, n and p respectively, but also the electron and hole drift mobilities, μ_e and μ_h . In the first approximation, the Hall coefficient can be calculated in the following way. Both charge carriers experience a Lorentz force in the same direction since they are drifting in the opposite directions but of course have opposite signs as illustrated in Fig. 2.14. Thus, both holes and electrons accumulate near the bottom surface. The magnitude of the Lorentz force, however, will be different since the drift mobilities and hence drift velocities will be different. Once equilibrium is reached, there should be no current flowing in the y -direction as we have an open circuit. The latter physical arguments lead to the following Hall coefficient [2.5],

$$R_H = \frac{p\mu_h^2 - n\mu_e^2}{e(p\mu_h + n\mu_e)^2}$$

or
$$R_H = \frac{p - nb^2}{e(p + nb)^2}, \quad (2.61)$$

where $b = \mu_e/\mu_h$.

It is clear that the Hall coefficient depends on both the drift mobility ratio and the concentrations of holes and electrons. For $p > nb^2$, R_H will be positive and for $p < nb^2$, it will be negative. Note that the carrier concentration is not zero when the Hall coefficient is zero but rather $n/p = (\mu_h/\mu_e)^2$. As an example, Fig. 2.15 shows the dependence of Hall coefficient vs. electron concentration for a single crystal silicon. The calculations are based on (2.61) and the law of mass action

$$np = n_i^2, \quad (2.62)$$

where n_i is the electron concentration in the intrinsic semiconductor.

In the case of *monopolar conduction*, e.g. conduction in metals or in doped semiconductors, (2.61) reduces to

$$R_H = -\frac{1}{en}, \quad (\text{for } n \gg p) \quad (2.63)$$

or

$$R_H = \frac{1}{ep}, \quad (\text{for } p \gg n). \quad (2.64)$$

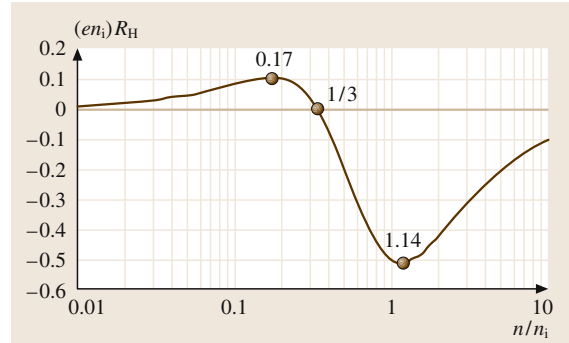


Fig. 2.15 Normalized Hall coefficient versus normalized electron concentration plot for single-crystal silicon. The values 0.17, 1.14 and 0.33 shown are the n/n_i values when the magnitude R reaches maxima and zero respectively. In single-crystal silicon, $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$, $\mu_e = 1350 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $\mu_h = 450 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$

Therefore, (considering as an example a n -type semiconductor where $\sigma = ne\mu_n$) one can write

$$\mu_H = \frac{\sigma}{ne} = -\sigma R_H \quad (2.65)$$

which provides a simple expression for determining the electron mobility known as the *Hall mobility*. Note, however, that the Hall mobility may differ from the drift mobility discussed in the previous sections. The difference arises from the carriers in a semiconductor having a distribution of energies. An average is used to describe the effect of carriers occupying different allowed states. (This is distinct from the earlier discussions of mobility where it was assumed that all the carriers have the same mean free time between collisions.) To include this, it is necessary to use a formal analysis based on the Boltzmann transport equation, as discussed in Sect. 2.6. If we express the averaging of the carriers with energy E and distribution function $f(E)$, we can write the energy-averaged $\tau(E)$ (i. e., $\langle \tau \rangle$) as:

$$\langle \tau \rangle = \frac{\int \tau(E) f(E) dE}{\int f(E) dE} \quad (2.66)$$

and the energy-averaged τ^2 (i. e., $\langle \tau^2 \rangle$) as:

$$\langle \tau^2 \rangle = \frac{\int \tau^2(E) f(E) dE}{\int f(E) dE}. \quad (2.67)$$

The rigorous analysis [2.12] shows that the Hall mobility μ_H in terms of the drift mobility μ_d is

$$\mu_H = r_H \times \mu_d, \quad (2.68)$$

where r_H is the *Hall factor* given by the ratio

$$r_H = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2}. \quad (2.69)$$

2.10 High Electric Field Transport

The previous sections focused on carrier transport in weak electric fields where the energy gained by carriers from the field is lost to the lattice through collisions with phonons or ionized impurities. However at higher fields, the efficiency of collision mechanisms diminishes and the carrier system contains more energy than the lattice. The carriers are then called *hot* with effective temperatures T_e for electrons and T_h for holes. In this case, the drift velocity no longer obeys Ohm's law, and becomes non-linear in the applied field with a clear tendency to saturation due to the appearance of a new dissipation mechanism involving optical phonon generation. Figure 2.16 shows the drift velocity saturation for both electrons and holes in Si and electrons in GaAs; GaAs shows a region of electron velocity overshoot and then negative differential resistivity due to inter-valley scattering, i. e., the transfer of electrons from the Γ minimum to the L conduction band minimum.

Solving the Boltzmann transport equation by analogy with (2.6, 2.7) the mobility may be defined as

$$\mu = \frac{e \langle \tau(T_e) \rangle}{m^*}, \quad (2.70)$$

where e is electron charge and m_e^* is the effective mass and $\langle \tau(T_e) \rangle$ is the mean free time which now strongly de-

pends on T_e . Therefore, the high-field mobility is related to the low field mobility μ_0 by

$$\mu = \mu_0 \left(\frac{T_e}{T} \right)^\beta, \quad (2.71)$$

where β depends on electric field and scattering mechanisms. Thus, in order to determine the dependence of the mobility on electric field, the dependence of the effective carrier temperature on field is required. This may be found by using the time-dependent Boltzmann transport equation. Suppose that F is the field, then for non-degenerate conditions with $T_e \gg T$

$$\frac{T_e}{T} \propto F^{-\frac{2}{2\beta-1}} \quad (2.72)$$

and hence

$$\mu(E) \propto \mu_0 F^{-\frac{2}{2\beta-1}}. \quad (2.73)$$

For acoustic phonon scattering, $\beta = -3/2$ and the drift mobility shows no saturation, increasing with field as $F^{1/4}$. Saturation in the drift velocity may be achieved only when $\beta \rightarrow \infty$ due to optical phonon scattering where large energy changes are involved. The saturation velocity v_s (related to the saturation mobility as $v_s = F\mu_s$) may be calculated using the energy and momentum rate equations for optical phonon scattering:

$$\frac{d\langle E \rangle}{dt} = eFv_s - \frac{E_{op}}{\tau_e} \quad (2.74)$$

$$\frac{d\langle m^* v_s \rangle}{dt} = eF - \frac{m^* v_s}{\tau_m} \quad (2.75)$$

where E_{op} is the optical phonon energy, τ_e and τ_m are the energy and momentum relaxation times, respectively. At steady state and for not extremely high fields, one may assume that $\tau_e \approx \tau_m$. (It is worth noting that at the high-electric fields $\tau_e > \tau_m$ and may lead to the appearance of avalanche, as discussed in Sect. 2.11). Therefore, the solution of (2.74, 2.75) is

$$v_s = \left(\frac{E_{op}}{m^*} \right)^{1/2}. \quad (2.76)$$

in agreement with the experimental values shown in Fig. 2.16.

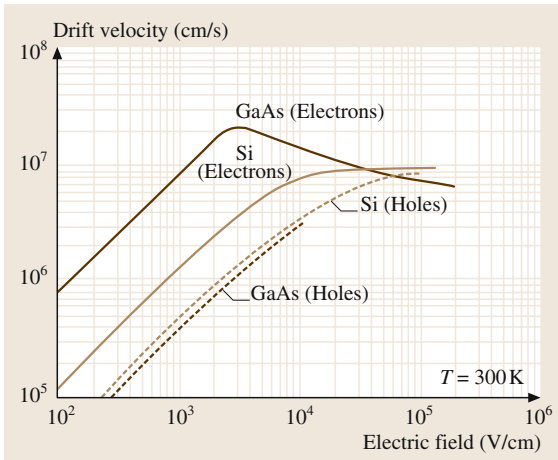


Fig. 2.16 Dependence of carrier drift velocity on electric field for GaAs and Si [2.53]

2.11 Avalanche

At very high electric field (in the range 2×10^5 V/cm or larger) a new possibility appears: a carrier may have kinetic energy in excess of the binding energy of a valence electron to its parent atom. In colliding with an atom, such a carrier can break the covalent bond and produce an electron-hole pair. This process is called *impact ionization* and is characterized by the *impact ionization coefficient* α (α_e for electrons and α_h for holes). The released electrons and holes may, in turn, impact ionize more atoms producing new electron-hole pairs. This process is called *avalanche* and is characterized by the *multiplication factor*, M which is the ratio of number of collected carriers to the number of initially injected carriers.

The field dependence of the impact ionization coefficient, at least over the limited fields where avalanche

is observed, is usually modeled by experimentalists by using

$$\alpha = A \exp \left[- \left(\frac{B}{F} \right)^n \right], \quad (2.77)$$

where F is the field A , B , n are constants that depend on the semiconductor material properties such as the E - k electronic structure, phonon energies and spectra, scattering mechanisms and so on. The constant B has been semiquantitatively argued to depend on E_g/τ where E_g is the bandgap and $1/\tau$ is the phonon scattering rate; higher bandgap semiconductors tend to have steeper slopes on $\log \alpha$ vs. $1/F$ plots and the $\log \alpha$ vs. $1/F$ curve tends to shift to higher fields. Typically, it has been found that $n \approx 1$ at low fields and $n \approx 2$ for high fields. Figure 2.17 shows experimental data for a variety of materials over a wide range of electric fields.

The origin of (2.77) in its simple $n = 1$ form lies in Shockley's [2.64] *lucky electron* model. When a carrier moves a distance z down-stream (along the field) without being scattered, it gains an energy eFz . An unlucky carrier is scattered so frequently that its eFz never reaches the threshold ionization energy E_I for impact ionization. On the other hand, a *lucky electron* is a ballistic electron that avoids scattering for a substantial distance, and hence is able to build its eFz to reach E_I and thereby cause impact ionization. If λ is the mean free path of collisions, then Shockley's model gives

$$\alpha = \frac{eF}{E_I} \exp \left(- \frac{E_I}{e\lambda F} \right). \quad (2.78)$$

The main problem with the Shockley model is that there are just not enough ballistic electrons to cause sufficient impact ionizations to explain the experiments. A better model was developed by Baraff [2.65] who numerically solved the Boltzmann transport equation for a simple parabolic band and energy independent mean free path to provide a relationship between α and F in terms of four parameters, that is, threshold energy for impact ionization, mean free path associated with ionizations, optical phonon energy, and mean free path for optical phonon scattering. Baraff's theory served experimentalists quite well in terms of comparing their results even though the model was not intuitive and was limited in terms of its assumptions and applicability to real semiconductors.

Impact ionization theory in crystalline solids only reached an acceptable level of confidence and understanding in the 1980s and 1990s with the development of the lucky-drift model by Ridley [2.66] and its extension

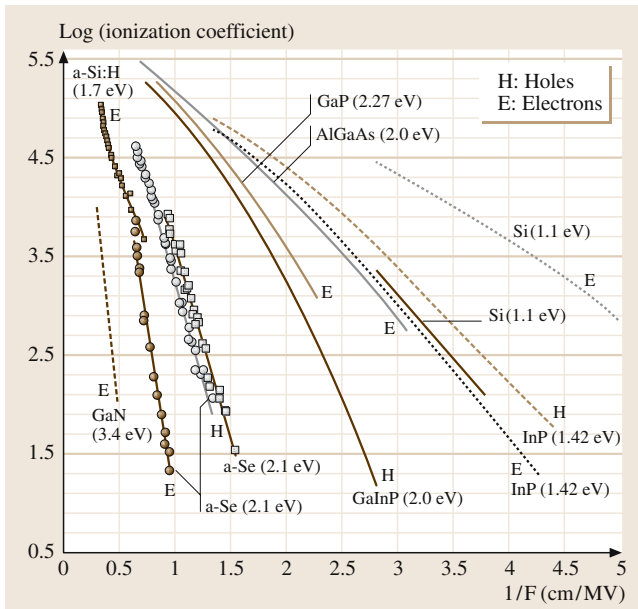


Fig. 2.17 Semilogarithmic plot of the dependence of the impact ionization coefficient on the reciprocal field for not only a-Se and a-Si:H, but also for various crystalline semiconductors for comparison; H indicates holes and E electrons. The y-axis is a base-10 logarithm of α in which α is in $1/\text{cm}$. Data for a-Si:H from [2.54]; a-Se electrons and holes from [2.55]; a-Se holes from [2.56], from which α_h was obtained by reanalyzing their multiplication data. Data for crystalline semiconductors are for GaP [2.57]; GaInP ($\text{Ga}_{0.52}\text{In}_{0.48}\text{P}$) [2.58]; $\text{Al}_{0.60}\text{Ga}_{0.40}\text{As}$, [2.59]; InP, [2.60]; Si, [2.61]; GaN (calculation only), [2.62]. Solid circle: Electrons in a-Se [2.55]; Solid square: electrons in a-Si:H [2.54]; Open circle: holes in a-Se [2.55]; Open square: holes in a-Se [2.63]

by Burt [2.67], and Mackenzie and Burt [2.68]. The latter major advancement in the theory appeared as the *lucky drift (LD) model*, and it was based on the realization that at high fields, hot electrons do not relax momentum and energy at the same rates. Momentum relaxation rate is much faster than the energy relaxation rate. An electron can drift, being scattered by phonons, and have its momentum relaxed, which controls the drift velocity, but it can still gain energy during this drift. Stated differently, the mean free path λ_E for energy relaxation is much longer than the mean free path λ for momentum relaxation.

In the Mackenzie and Burt [2.68] version of the LD model, the probability $P(E)$ that a carrier attains an energy E is given by,

$$P(E) = \exp\left(-\int_0^E \frac{dE'}{eF\lambda(E')}\right) + \int_0^E \frac{dE_1}{eF\lambda(E_1)} \times \exp\left(-\int_0^{E_1} \frac{dE'}{eF\lambda(E')}\right) \times \exp\left(-\int_{E_1}^E \frac{dE'}{eF\lambda_E(E')}\right), \quad (2.79)$$

where as mentioned above λ is the mean free path associated with momentum relaxing collisions and λ_E is the mean energy relaxation length associated with the energy relaxing collisions. The first term is the Shockley lucky electron probability, i.e. the electron moves ballistically to energy E . The second term is the lucky drift probability term which is composed of the following: the electron first moves ballistically to some intermediate energy E_1 ($0 < E_1 < E$) from where it begins its lucky drift to energy E_1 ; hence the integration over all possible E_1 . The impact ionization coefficient can then

readily be evaluated from

$$\alpha = \frac{eFP(E_1)}{E_1 \int_0^{E_1} P(E) dE}. \quad (2.80)$$

The model above is based on a hard threshold ionization energy E_1 , that is, when a carrier attains the energy E_1 , ionization ensues. The model has been further refined by the inclusion of soft threshold energies which represent the fact the ionization does not occur immediately when the carrier attains the energy E_1 , and the carrier drifts further to gain more energy than E_1 before impact ionization [2.69–71].

Assuming λ and λ_E are energy independent, which would be the case for a single parabolic band in the crystalline state, (2.79) and (2.80) can be solved analytically to obtain

$$\alpha = \frac{1}{\lambda} \times \frac{\frac{\lambda}{\lambda_E} \exp\left(\frac{-E_1}{eF\lambda_E}\right) + \left(\frac{\lambda}{\lambda_E}\right)^2 \exp\left(\frac{-E_1}{eF\lambda}\right)}{1 - \exp\left(\frac{-E_1}{eF\lambda_E}\right) - \left(\frac{\lambda}{\lambda_E}\right)^2 \left[1 - \exp\left(\frac{-E_1}{eF\lambda}\right)\right]}. \quad (2.81)$$

For $\lambda_E > \lambda$, and in the “low field region”, where typically $(\alpha\lambda) < 10^{-1}$, or $x = E_1/eF\lambda > 10$, (2.81) leads to a simple expression for α ,

$$\alpha = \left(\frac{1}{\lambda_E}\right) \exp\left(-\frac{E_1}{eF\lambda_E}\right). \quad (2.82)$$

For crystalline semiconductors, one typically also assumes that λ_E depends on the field F , λ and the optical phonon energy $\hbar\omega$ as

$$\lambda_E = \frac{eF\lambda^2}{2\hbar\omega} \coth\left(\frac{\hbar\omega}{2kT}\right). \quad (2.83)$$

As the field increases, λ_E eventually exceeds λ , and allows lucky drift to operate and the LD carriers to reach the ionization energy.

It is worth noting that the model of lucky drift is successfully used not only for crystalline semiconductors but to amorphous semiconductors [2.72].

2.12 Two-Dimensional Electron Gas

Heterostructures offer the ability to spatially engineer the potential in which carriers move. In such structures having layers deposited in the z -direction, when the width of a region with confining potential $t_z < \lambda_{dB}$, the de Broglie electron wavelength, electron states become stationary states in that direction, retaining Bloch wave character in the other two directions (i.e., x - and

y -directions), and is hence termed a *2-D electron gas (2DEG)*. These structures are notable for their extremely high carrier mobility.

High mobility structures are formed by selectively doping the wide bandgap material behind an initially undoped spacer region of width d as shown in Fig. 2.18a. Ionization and charge transfer leads to carrier build-up

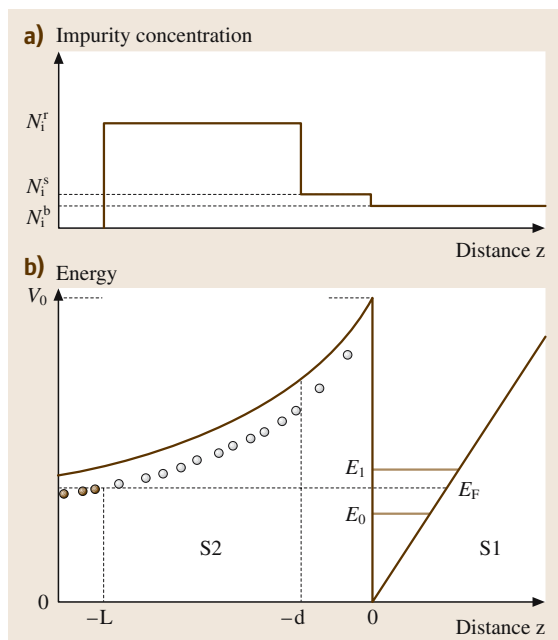


Fig. 2.18 (a) Doping profile for a selectively doped 2DEG heterostructure. The S1 is a narrow bandgap material adjacent to the heterointerface. The S2 is selectively doped wide-bandgap material including an initially undoped spacer region of width d . (b) Conduction energy band structure for a selectively doped 2DEG heterostructure. The equilibrium band bending (through Poisson's equation) in the well region results from the equalization of the ionized donor concentration in the wide-bandgap material and the 2DEG concentration adjacent to the heterointerface. When the associated interfacial field is sufficiently strong, carriers are confined within λ_{dB} and electron states are quantized into the sub-bands E_0 and E_1 [2.73]

in the low potential region of narrow bandgap material adjacent to the hetero-interface. The equilibrium band bending (i.e., through Poisson's equation) in the well region, as shown in Fig. 2.18b, results from the equalization of ionized donor concentration in the wide bandgap material and 2DEG concentration adjacent to the heterointerface. When the associated interfacial field is sufficiently strong, carriers are confined within λ_{dB} and electron states are quantized into sub-bands (i.e., E_0 and E_1), as shown in Fig. 2.18b.

Figure 2.19 shows the contributions of component scattering mechanisms to the low temperature mobility of a 2DEG formed at a $\text{Ga}_{0.70}\text{Al}_{0.3}\text{As}$ -GaAs heterointerface, as a function of the electron gas density. As for bulk samples, the most important mechanism limiting

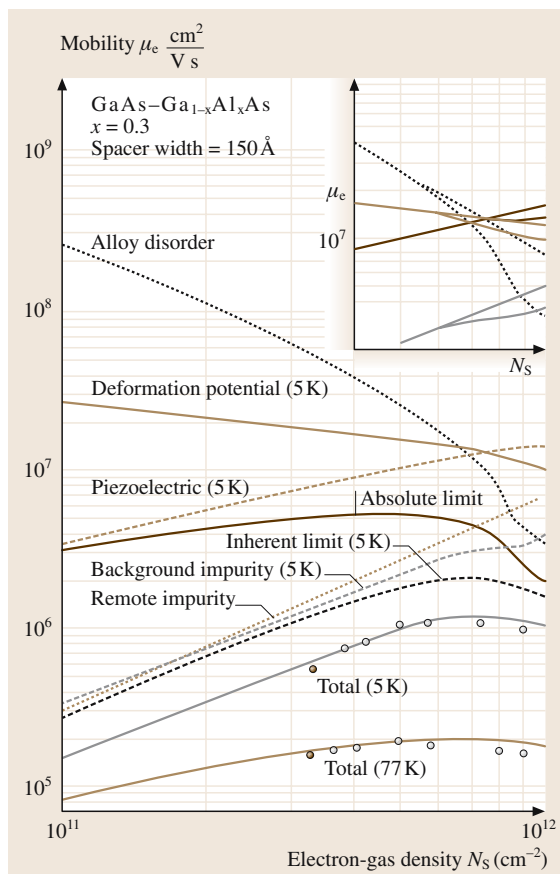


Fig. 2.19 Dependence of 2DEG electron mobility on carrier concentration for a $\text{Ga}_{0.70}\text{Al}_{0.3}\text{As}$ -GaAs heterostructure [2.73]

the low temperature mobility is ionized impurity scattering, except at high electron densities, where so-called *alloy disorder scattering* is significant. Ionized impurity scattering may be further broken down into scattering from ionized impurities that are with the GaAs quantum well, known as *background impurities*, those beyond the spacer region, termed *remote impurity scattering*. For high purity growth, the unintentional background impurity concentration can be kept to very low limits and impurity scattering based mobility values are then dictated by remote impurity scattering. Since carriers in the well are only weakly scattered by the tail field of these remote Coulomb centers, the mobility of such 2DEG systems can be orders of magnitude higher than bulk samples. The temperature dependence of the electron mobility for such a system is shown in Fig. 2.20. Notice how, similar to bulk sample, increasing temperature

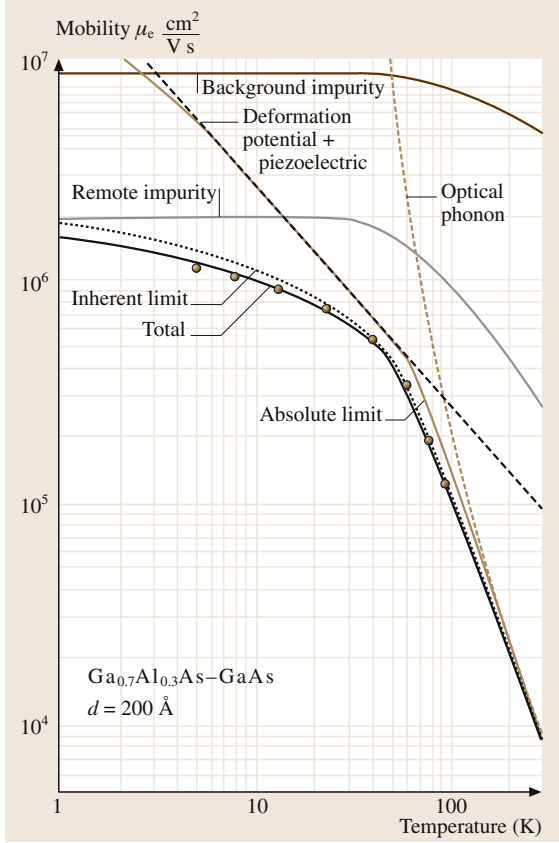


Fig. 2.20 Dependence of 2DEG electron mobility on temperature for a $\text{Ga}_{0.70}\text{Al}_{0.3}\text{As-GaAs}$ heterostructure [2.73]

increases the phonon population such that for the example shown, above about 100 K, polar optical phonon scattering controls the mobility.

Figure 2.21 shows the dependence of the 2DEG mobility on the spacer width. Two competing factors are active – for narrow spacer widths, the transfer efficiency of carriers to the GaAs well is high and so a lower remote doping concentration is sufficient to provide for a given constant 2DEG concentration, but since the Coulomb

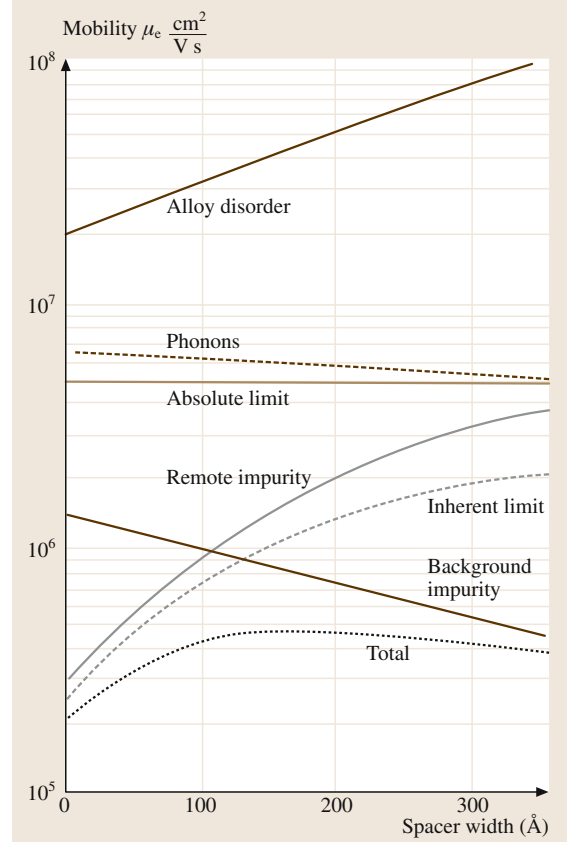


Fig. 2.21 Dependence of 2DEG electron mobility on spacer width for a $\text{Ga}_{0.70}\text{Al}_{0.3}\text{As-GaAs}$ heterostructure [2.73]

scatters are so close to the 2DEG, they scatter very efficiently and limit the mobility. On the other hand, for large spacer widths, carrier transfer efficiency is quite poor requiring higher remote doping to supply the given 2DEG concentration; however, being relatively far away each of the scattering centers are less effective at lowering the mobility, but given their high concentration, the net effect is still a decrease in the mobility at large spacer widths as seen in the figure.

2.13 One Dimensional Conductance

In the case where carriers are confined within regions of width L_x , $L_y < \lambda_{dB}$ in two directions x and y , respectively, the electron energy in those two directions become quantized with quantum numbers n_x and n_y , respectively. In the third direction z , electrons travel as Bloch waves with energy that may be approximated by

$\hbar^2 k_z^2 / 2m^*$ giving an expression for the total energy E of the so-called *I-D electron system* or *quantum wire* as:

$$E(n_x, n_y, n_z) = \frac{\hbar^2 \pi^2}{2m^*} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} \right) + \frac{\hbar^2 k_z^2}{2m^*}. \quad (2.84)$$

The associated electron wavefunctions are:

$$\Psi(x, y, z) = \frac{1}{2\sqrt{L_x L_y L_z}} \sin\left(\frac{n_x \pi x}{L_x}\right) \times \sin\left(\frac{n_y \pi y}{L_y}\right) e^{ik_z z}. \quad (2.85)$$

Using these equations, one can readily derive an expression for the density of states per unit energy range:

$$\begin{aligned} \text{DOS} &= 2 \times 2 \left(\frac{L_z}{2\pi}\right) (\nabla_{k_z} E)^{-1} \\ &= \frac{2L_z}{h} \sqrt{\frac{m^*}{2(E - E_{n_x, n_y})}}. \end{aligned} \quad (2.86)$$

In order to evaluate the conductance of this quantum wire, consider the influence of a weak applied potential V . Similar to the case for bulk transport the applied field displaces the *Fermi surface* and results in a change in the electron wave-vector from k_0 (i. e., with no applied potential) to k_V (i. e., when the potential is applied). When V is small compared with the electron energy:

$$k_0 = \sqrt{\frac{2m^*(E - E_{n_x, n_y})}{\hbar^2}}, \quad (2.87)$$

$$\begin{aligned} k_V &= k_0 \sqrt{1 + \frac{eV}{E - E_{n_x, n_y}}} \\ &\approx k_0 \left(1 + \frac{1}{2} \frac{eV}{E - E_{n_x, n_y}}\right). \end{aligned} \quad (2.88)$$

This leads to establishing a current density J in the wire

$$J = \frac{2e^2 (\text{DOS}) \sqrt{(E_F - E_{n_x, n_y})}}{\sqrt{2m^*}}. \quad (2.89)$$

Which may be simplified to the following expressions for J and the current flowing in the wire for a given quantum state $E\{n_x, n_y\}$, I

$$J_{n_x, n_y} = \frac{2e^2 V L_z}{h} \quad \text{and} \quad I_{n_x, n_y} = \frac{2e^2 V}{h}. \quad (2.90)$$

The expression for the conductance through one channel corresponding to a given quantum state $\{n_x, n_y\}$ is then given by

$$G_{n_x, n_y} = \frac{I_{n_x, n_y}}{V} = \frac{2e^2}{h}. \quad (2.91)$$

Notice how the conductance is quantized in units of e^2/h with each populated channel contributing equally to the conductance – moreover, this is a fundamental result, being independent of the material considered. In practice, deviations from this equation can occur (although generally less than 1%) owing to the finiteness of real nanowires and impurities in or near the channel, influencing the conductivity and even resulting in weak localization. Generally, unlike both bulk and 2DEG systems, ionized impurity scattering is suppressed in nanowires. The main reason for this is that an incident electron in a quantum state $\{n_x, n_y\}$ traveling along the wire with wave-vector $k_z\{n_x, n_y\}$, can not be elastically scattered into any states except those in a small region of k -space in the vicinity of $-k_z\{n_x, n_y\}$. Such a scattering event involves a large change in momentum of $\approx 2k_z\{n_x, n_y\}$ and thus, the probability of such events is very small. As a result, the mean free path and mobility of carriers in such quantum wires are substantially increased.

The nature of carrier transport in quantum wires depends on the wire dimensions (i. e., length L_{Wire} and diameter d_{Wire}) as compared with the carrier mean free path, l_{Carrier} . When $l_{\text{Carrier}} \gg L_{\text{Wire}}, d_{\text{Wire}}$ the only potential seen by the carriers is that associated with the wire walls, and carriers exhibit wavelike behavior, being guided through the wire as if it were a waveguide without any internal scattering. Conversely, if $d_{\text{Wire}} \ll \lambda_{\text{DeBroglie}}$, only a few energy states in the wire are active, and in the limit of an extremely small waveguide, only one state or channel is active, analogous to a single mode waveguide cavity – this case is termed *quantum ballistic transport*. In the limit, $l_{\text{Carrier}} \ll L_{\text{Wire}}, d_{\text{Wire}}$, scattering dominates transport throughout the wire – with numerous scattering events occurring before a carrier can traverse the wire or move far along its length. In such a case the transport is said to be *diffusive*. As discussed previously, ionized impurity and lattice scattering contribute to l_{Carrier} , with l_{Carrier} decreasing with increasing temperature due to phonon scattering. For strong impurity scattering, this may not occur until relatively high temperatures. In the intermediate case of $L_{\text{Wire}} \gg l_{\text{Carrier}} \gg d_{\text{Wire}}$ and where $d_{\text{Wire}} \ll \lambda_{\text{DeBroglie}}$ scattering is termed “mixed mode” and is often called *quasi-ballistic*.

2.14 The Quantum Hall Effect

The observation of, and first explanation for the Hall Effect in a 2DEG by von Klitzing et al. [2.74], won them a Nobel Prize. As shown in Fig. 2.22 the Hall re-

sistivity exhibits plateaus for integer values of h/e^2 , independent of any material dependent parameters. This discovery was later shown to be correct to a precision

of at least one part in 10^7 , enabling extremely accurate determinations of the fine structure constant (i.e., $\alpha = (\mu_0 c e^2 / 2h) \approx 1/137$) and a fundamental resistance standard to be established.

For the six point Hall geometry used (as shown in the insert to Fig. 2.18), one can define the *Hall resistivity* $\rho_{xy} = g_1 V_{H,y} / I_x$ and *longitudinal resistivity* $\rho_{xx} = g_2 V_{L,x} / I_x$ where g_1 and g_2 are geometric constants related to the sample geometry. These resistivities are related to corresponding conductivities through the conductivity and resistivity tensors

$$\begin{aligned}\rho_{xy} &= \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2}, \\ \rho_{xx} &= \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2}, \\ \rho_{yx} &= -\rho_{xy}, \\ \rho_{yy} &= \rho_{xx}.\end{aligned}\quad (2.92)$$

Starting from a classical equation of motion for electrons in an electric field E_x , magnetic field B_z , and defining the *cyclotron frequency* $\omega_c = eB_z/m^*$, the velocities perpendicular to the applied magnetic field can be deduced as

$$v_x = \frac{E_x}{B_x} \sin \omega_c t \quad \text{and} \quad v_y = \frac{E_x}{B_x} (\cos \omega_c t - 1) \quad (2.93)$$

with the time averaged velocities

$$\langle v_x \rangle = \frac{E_x}{B_x} \quad \text{and} \quad \langle v_y \rangle = \frac{E_x}{B_x} \quad (2.94)$$

leading to Hall and longitudinal resistivities of

$$\sigma_{xx} = \sigma_{yy} = 0 \quad \text{and} \quad \sigma_{xy} = -\sigma_{yx} = \frac{N_s e}{B_z}, \quad (2.95)$$

where N_s is the areal electron concentration. Below, a quantum approach is used to establish a relationship for the electron concentration N_s . Note that the motion of the electrons in the crossed fields are quantized with allowed levels termed *Landau levels* E_n :

$$E_n = \left(n + \frac{1}{2}\right) \hbar \omega_c + g^* \mu_B B_z + \varepsilon_z, \quad (2.96)$$

where n is the quantum number describing the particular Landau Level, g^* is the *Landé factor*, μ_B is the Bohr magneton and $g^* \mu_B B_z$ is the *spin magnetic energy*, and E_z is the energy associated with the z -motion of the carriers. xy -plane carrier motion is characterized by the cyclotron energy term E_{xy} ,

$$E_{xy} = \frac{\hbar^2 k_{xy}^2}{2m_e^*} = \left(n + \frac{1}{2}\right) \hbar \omega_c. \quad (2.97)$$

Following this description and noting that the motion of electrons in the xy -plane may be expressed in terms of wavefunctions of the *harmonic oscillator* using the *Landau gauge* of the *vector potential* $[0, xB_z, 0]$, we may write the density of states per unit area, DOS_A as:

$$\text{DOS}_A = \frac{m_e^* \omega_c L_x L_y}{2\pi \hbar}. \quad (2.98)$$

Since the degeneracy of each Landau level is one (i.e., since they are single spin states), this enables one to find N_s assuming Landau state filling up to the p^{th} level (for integer p):

$$N_s = \frac{m_e^* \omega_c p}{2\pi \hbar}. \quad (2.99)$$

Using the definition of the cyclotron frequency gives the final form

$$N_s = \frac{peB_z}{\hbar} \quad (2.100)$$

which may be used to rewrite the previous expressions for Hall and longitudinal resistivity

$$\rho_{xx} = \rho_{yy} = 0 \quad \text{and} \quad \rho_{xy} = -\rho_{yx} = \frac{h}{pe^2}. \quad (2.101)$$

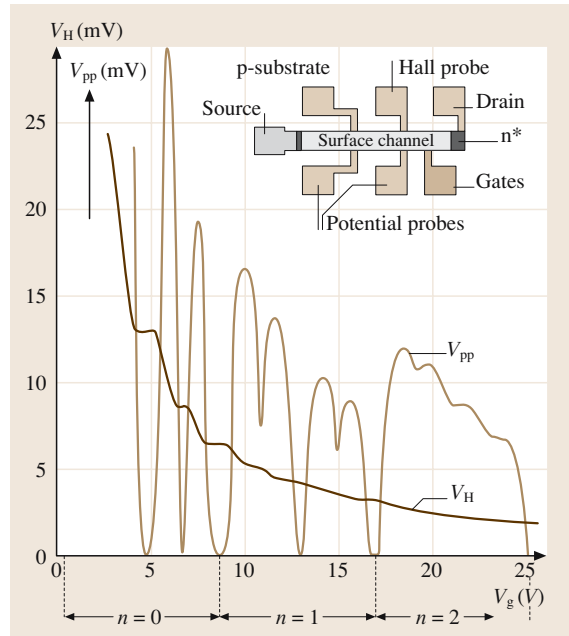


Fig. 2.22 Hall voltage V_H and voltage drop across electrodes V_{PP} as a function of gate voltage V_g at 1.5 K, when $B = 18$ T. Source-drain current is $1 \mu\text{A}$. Insert shows plan view of a device with length $400 \mu\text{m}$, width $50 \mu\text{m}$ and an interprobe separation of $130 \mu\text{m}$ [2.74]

Equation (2.101) shows that Hall resistivity is quantized in units of h/pe^2 whenever the Fermi energy lies between filled Landau levels. Consistent with observation, the result is independent of the semiconductor being studied. Although this model provides an excel-

lent basis for understanding experiments, understanding the details of the results (i.e., in particular the existence of a finite width for the Hall effect plateaus and zero longitudinal resistance dips) requires a more complete treatment involving so-called localized states.

References

- 2.1 P. L. Rossiter: *The Electrical Resistivity of Metals and Alloys* (Cambridge Univ. Press, Cambridge 1987)
- 2.2 J. S. Dugdale: *The Electrical Properties of Metals and Alloys* (Arnold, London 1977)
- 2.3 B. R. Nag: *Theory of Electrical Transport in Semiconductors* (Pergamon, Oxford 1972)
- 2.4 F. J. Blatt: *Physics of Electronic Conduction in Solids* (McGraw-Hill, New York 1968) Chap. 5, 6
- 2.5 S. O. Kasap: *Principles of Electronic Materials and Devices*, 3 edn. (McGraw-Hill, New York 2005)
- 2.6 G. T. Dyos, T. Farrell (Eds.): *Electrical Resistivity Handbook* (Peregrinus, London 1992)
- 2.7 D. G. Fink, D. Christiansen (Eds.): *Electronics Engineers' Handbook*, 2 edn. (McGraw-Hill, New York 1982) Section 6
- 2.8 L. Nordheim: Ann. Phys. **9**, 664 (1931)
- 2.9 J. K. Stanley: *Electrical and Magnetic Properties of Metals* (American Society for Metals, Metals Park 1963)
- 2.10 M. Hansen, K. Anderko: *Constitution of Binary Alloys*, 2 edn. (McGraw-Hill, New York 1985)
- 2.11 H. E. Ruda: J. Appl. Phys. **59**, 1220 (1986)
- 2.12 M. Lundstrom: *Fundamentals of Carrier Transport* (Cambridge Univ. Press, Cambridge 2000)
- 2.13 R. H. Bube: *Electronic Properties of Crystalline Solids* (Academic, New York 1974) Chap. 7
- 2.14 S. Riedel, J. Röber, T. Geßner: Microelectron. Eng., **33**, 165 (1997)
- 2.15 A. F. Mayadas, M. Shatzkes: Phys. Rev. B, **1**, 1382 (1970)
- 2.16 J.-W. Lim, K. Mimura, M. Isshiki: Appl. Surf. Sci. **217**, 95 (2003)
- 2.17 C. R. Tellier, C. R. Pichard, A. J. Tosser: J. Phys. F, **9**, 2377 (1979) (and references therein)
- 2.18 K. Fuchs: Proc. Camb. Philos. Soc., **34**, 100 (1938)
- 2.19 E. H. Sondheimer: Adv. Phys., **1**, 1 (1952)
- 2.20 H.-D. Liu, Y.-P. Zhao, G. Ramanath, S. P. Murarka, G.-C. Wang: Thin Solid Films **384**, 151 (2001)
- 2.21 R. Suri, A. P. Thakoor, K. L. Chopra: J. Appl. Phys., **46**, 2574 (1975)
- 2.22 R. H. Cornely, T. A. Ali: J. Appl. Phys., **49**, 4094 (1978)
- 2.23 J. S. Ahn, K. H. Kim, T. W. Noh, D. H. Riu, K. H. Boo, H. E. Kim: Phys. Rev. B, **52**, 15244 (1995)
- 2.24 R. J. Gehr, G. L. Fisher, R. W. Boyd: J. Opt. Soc. Am. B, **14**, 2310 (1997)
- 2.25 D. E. Aspnes, J. B. Theeten, F. Hottier: Phys. Rev. B, **20**, 3292 (1979)
- 2.26 Z. Yin, F. W. Smith: Phys. Rev. B, **42**, 3666 (1990)
- 2.27 M. F. MacMillan, R. P. Devaty, W. J. Choyke, D. R. Goldstein, J. E. Spanier, A. D. Kurtz: J. Appl. Phys., **80**, 2412 (1996)
- 2.28 C. Ganter, W. Schirmacher: Phys. Status Solidi B, **218**, 71 (2000)
- 2.29 R. Stognienko, Th. Henning, V. Ossenkopf.: Astron. Astrophys. **296**, 797 (1995)
- 2.30 A. G. Rojo, H. E. Roman: Phys. Rev. B, **37**, 3696 (1988)
- 2.31 J. A. Reynolds, J. M. Hough: Proc. Phys. Soc., **70**, 769 (1957)
- 2.32 R. Clausius: *Die Mechanische Wärmetheorie*, Vol. 2 (Wieveg, Braunschweig 1879)
- 2.33 L. Lorenz: Ann. Phys. Lpz., **11**, 70 (1880)
- 2.34 O. F. Mosotti: Mem. Math. Fisica Modena II, **24**, 49 (1850)
- 2.35 V. I. Odelevskii: Zh. Tekh. Fiz., **6**, 667 (1950)
- 2.36 Lord Rayleigh: Philos. Mag., **34**, 481 (1892)
- 2.37 K. W. Wagner: Arch. Electrochem., **2**, 371 (1914)
- 2.38 D. A. G. Bruggeman: Ann. Phys. Lpz. **24**, 636 (1935)
- 2.39 C. J. F. Bottcher: Rec. Trav. Chim. Pays-Bas **64**, 47 (1945)
- 2.40 H. Fricke: Phys. Rev. **24**, 575 (1924)
- 2.41 D. Polder, J. M. Van Santen: Physica **12**, 257 (1946)
- 2.42 W. Niesel: Ann. Phys. Lpz. **10**, 336 (1952)
- 2.43 J. A. Stratton: *Electromagnetic Theory* (McGraw-Hill, New York 1941)
- 2.44 O. Wiener: Abh. Sachs. Ges. Akad. Wiss. Math. Phys. **32**, 509 (1912)
- 2.45 L. Silberstein: Ann. Phys. Lpz. **56**, 661 (1895)
- 2.46 O. Wiener: Abh. Sachs. Ges. Akad. Wiss. Math. Phys. **32**, 509 (1912)
- 2.47 R. W. Sillars: J. Inst. Elect. Eng. **80**, 378 (1937)
- 2.48 F. Ollendorf: Arch. Electrochem. **25**, 436 (1931)
- 2.49 J. C. M. Maxwell-Garnett: Phil. Trans. R. Soc. Lond. **203**, 385 (1904)
- 2.50 H. Looyenga: Physica **31**, 401 (1965)
- 2.51 J. Monecke: J. Phys. Condens. Mat. **6**, 907 (1994)
- 2.52 C. F. Bohren, D. R. Huffman: *Absorption and Scattering of Light by Small Particles* (Wiley, New York 1983)
- 2.53 P. Y. Yu, M. Cardona: *Fundamentals of Semiconductors* (Springer, Berlin, Heidelberg 1996)

- 2.54 M. Akiyama, M. Hanada, H. Takao, K. Sawada, M. Ishida: Jpn. J. Appl. Phys **41**, 2552 (2002)
- 2.55 K. Tsuji, Y. Takasaki, T. Hirai, K. Taketoshi: J. Non-Cryst. Solids **14**, 94 (1989)
- 2.56 G. Juska, K. Arlauskas: Phys. Status Solidi **77**, 387 (1983)
- 2.57 R. A. Logan, H. G. White: J. Appl. Phys. **36**, 3945 (1965)
- 2.58 R. Ghin, J. P. R. David, S. A. Plimmer, M. Hopkinson, G. J. Rees, D. C. Herbert, D. R. Wight: IEEE Trans. Electron Dev. **ED45**, 2096 (1998)
- 2.59 S. A. Plimmer, J. P. R. David, R. Grey, G. J. Rees: IEEE Trans. Electron Dev. **ED47**, 21089 (2000)
- 2.60 L. W. Cook, G. E. Bulman, G. E. Stillma: Appl. Phys. Lett. **40**, 589 (1982)
- 2.61 C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleimack, W. Wiegmann: Phys. Rev. **134**, B766 (1964)
- 2.62 C. Bulutay: Semicond. Sci. Technol. **17**, L59 (2002)
- 2.63 G. Juska, K. Arlauskas: Phys. Status Solidi **59**, 389 (1980)
- 2.64 W. Shockley: Solid State Electron. **2**, 35 (1961)
- 2.65 G. A. Baraff: Phys. Rev. **128**, 2507 (1962)
- 2.66 B. K. Ridley: J. Phys. C **16**, 4733 (1983)
- 2.67 M. G. Burt: J. Phys. C **18**, L477 (1985)
- 2.68 S. MacKenzie, M. G. Burt: Semicond. Sci. Technol. **2**, 275 (1987)
- 2.69 B. K. Ridley: Semicond. Sci. Technol. **2**, 116 (1987)
- 2.70 J. S. Marsland: Solid State Electron. **30**, 125 (1987)
- 2.71 J. S. Marsland: Semicond. Sci. Technol. **5**, 177 (1990)
- 2.72 S. O. Kasap, J. A. Rowlands, S. D. Baranovskii, K. Tan-ioka: J. Appl. Phys. **96**, 2037 (2004)
- 2.73 W. Walukiewicz, H. E. Ruda, J. Lagowski, H. C. Gatos: Phys. Rev. B **30**, 4571 (1984)
- 2.74 K. V. Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45**, 494 (1980)

3. Optical Properties of Electronic Materials: Fundamentals and Characterization

Light interacts with materials in a variety of ways; this chapter focuses on refraction and absorption. Refraction is characterized by a material's refractive index. We discuss some of the most useful models for the frequency dependence of the refractive index, such as those due to Cauchy, Sellmeier, Gladstone–Dale, and Wemple–Di Dominico. Examples are given of the applicability of the models to actual materials. We present various mechanisms of light absorption, including absorption by free carriers, phonons, excitons and impurities. Special attention is paid to fundamental and excitonic absorption in disordered semiconductors and to absorption by rare-earth, trivalent ions due to their importance to modern photonics. We also discuss the effect of an external electric field on absorption, and the Faraday effect. Practical techniques for determining the optical parameters of thin films are outlined. Finally, we present a short technical classification of optical glasses and materials.

3.1	Optical Constants	47
3.1.1	Refractive Index and Extinction Coefficient	47
3.1.2	Kramers–Kronig Relations	49
3.2	Refractive Index	50
3.2.1	Cauchy Dispersion Equation	50
3.2.2	Sellmeier Dispersion Equation	51
3.2.3	Gladstone–Dale Formula	51
3.2.4	Wemple–Di Dominico Dispersion Relation	52
3.2.5	Group Index (N)	53
3.3	Optical Absorption	53
3.3.1	Lattice or Reststrahlen Absorption and Infrared Reflection	54
3.3.2	Free Carrier Absorption (FCA)	55
3.3.3	Band-to-Band or Fundamental Absorption	57
3.3.4	Exciton Absorption	63
3.3.5	Impurity Absorption	66
3.3.6	Effects of External Fields	69
3.4	Thin Film Optics	70
3.4.1	Swanepoel's Analysis of Optical Transmission Spectra ...	71
3.4.2	Ellipsometry	72
3.5	Optical Materials	74
3.5.1	Abbe Number or Constringence ...	74
3.5.2	Optical Materials	74
3.5.3	Optical Glasses	76
	References	76

3.1 Optical Constants

The changes that light undergoes upon interacting with a particular substance are known as the optical properties of that substance. These optical properties are influenced by the macroscopic and microscopic properties of the substance, such as the nature of its surface and its electronic structure. Since it is usually far easier to detect the way a substance modifies light than to investigate its macroscopic and microscopic properties directly, the optical properties of a substance are often used to probe other properties of the material. There are many optical properties, including the most well known: reflection, refraction, transmission and absorption. Many of these optical properties are associated with important optical constants, such as the refractive index and the extinction

coefficient. In this section we review these optical constants, such as the refractive index and the extinction coefficient. Books by *Adachi* [3.1], *Fox* [3.2] and *Simmons and Porter* [3.3] are highly recommended. In addition, *Adachi* also discusses the optical properties of III–V compounds in this handbook.

3.1.1 Refractive Index and Extinction Coefficient

The refractive index n of an optical or dielectric medium is the ratio of the velocity of light c in vacuum to its velocity v in the medium; $n = c/v$. Using this and Maxwell's equations, one obtains the well-known

formula for the refractive index of a substance as $n = \sqrt{\varepsilon_r \mu_r}$, where ε_r is the static constant or relative permittivity and μ_r is the magnetic permeability. As $\mu_r = 1$ for nonmagnetic substances, one gets $n = \sqrt{\varepsilon_r}$, which is very useful for relating the dielectric properties of a material to its optical properties at any particular frequency of interest. Since ε_r depends on the wavelength of the light, the refractive index depends on it too, which is called dispersion. In addition to dispersion, an electromagnetic wave propagating through a lossy medium (one that absorbs or scatters radiation passing through it) experiences attenuation, which means that it loses its energy due to various loss mechanisms such as the generation of phonons (lattice waves), photogeneration, free carrier absorption and scattering. In such materials, the refractive index is a complex function of the frequency of the light wave.

The complex refractive index N , with real part n and imaginary part K (called the extinction coefficient), is related to the complex relative permittivity, $\varepsilon_r = \varepsilon'_r - i\varepsilon''_r$ by:

$$N = n - iK = \sqrt{\varepsilon_r} = \sqrt{\varepsilon'_r - i\varepsilon''_r} \quad (3.1)$$

where ε'_r and ε''_r are, respectively, the real and imaginary parts of ε_r . Equation (3.1) gives:

$$n^2 - K^2 = \varepsilon'_r \quad \text{and} \quad 2nK = \varepsilon''_r. \quad (3.2)$$

The optical constants n and K can be determined by measuring the reflectance from the surface of a material as a function of polarization and the angle of incidence. For normal incidence, the reflection coefficient r is obtained as:

$$r = \frac{1 - N}{1 + N} = \frac{1 - n + iK}{1 + n - iK}. \quad (3.3)$$

The reflectance R is then defined by:

$$R = |r|^2 = \left| \frac{1 - n + iK}{1 + n - iK} \right|^2 = \frac{(1 - n)^2 + K^2}{(1 + n)^2 + K^2}. \quad (3.4)$$

Notice that whenever K is large, for example over a range of wavelengths, the absorption is strong, and the reflectance is almost unity. The light is then reflected, and any light in the medium is highly attenuated.

Optical properties of materials are typically presented by showing either the frequency dependence (dispersion relation) of n and K or ε'_r and ε''_r . An intuitive guide to explaining dispersion in insulators is based on a single oscillator model in which the electric field in the light induces forced dipole oscillations in the material (displaces the electron shells to oscillate about the nucleus) with a single resonant frequency

ω_0 [3.1]. The frequency dependences of ε'_r and ε''_r are then obtained as:

$$\varepsilon'_r = 1 + \frac{N_{\text{at}}}{\varepsilon_0} \alpha'_e \quad \text{and} \quad \varepsilon''_r = 1 + \frac{N_{\text{at}}}{\varepsilon_0} \alpha''_e, \quad (3.5)$$

where N_{at} is the number of atoms per unit volume, ε_0 is the vacuum permittivity, and α'_e and α''_e are the real and imaginary parts of the electronic polarizability, given respectively by:

$$\alpha'_e = \alpha_{e0} \frac{1 - (\omega/\omega_0)^2}{[1 - (\omega/\omega_0)^2]^2 + (\gamma/\omega_0)^2 (\omega/\omega_0)^2}, \quad (3.6a)$$

and

$$\alpha''_e = \alpha_{e0} \frac{(\gamma/\omega_0)(\omega/\omega_0)}{[1 - (\omega/\omega_0)^2]^2 + (\gamma/\omega_0)^2 (\omega/\omega_0)^2}, \quad (3.6b)$$

where α_{e0} is the dc polarizability corresponding to $\omega = 0$ and γ is the loss coefficient. Using (3.1), (3.2) and (3.5), it is then possible to study (3.6a, 6b), the frequency dependence of n and K . Figure 3.1a shows the dependence of n and K on the normalized frequency ω/ω_0 for a simple single electronic dipole oscillator of resonance frequency ω_0 . It is seen that n and K peak close to $\omega = \omega_0$. If a material has a $\varepsilon''_r \gg \varepsilon'_r$, then $\varepsilon_r \approx -i\varepsilon''_r$ and $n = K \approx \sqrt{\varepsilon''_r/2}$ is obtained from (3.2). Figure 3.1b shows the dependence of the reflectance R on the frequency. It is observed that R reaches its

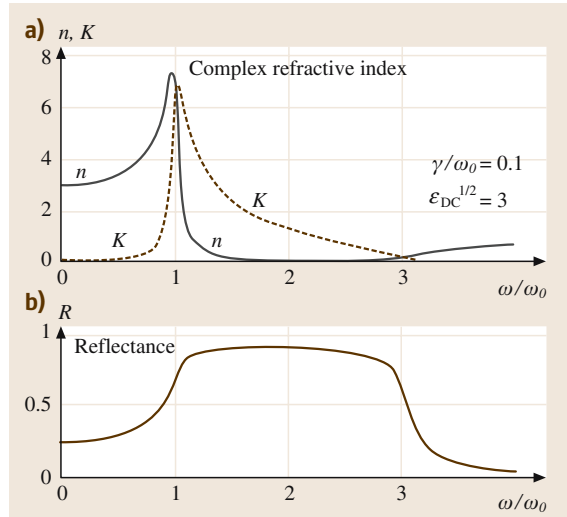


Fig. 3.1a,b The dipole oscillator model. (a) Refractive index and extinction coefficient versus normalized frequency. (b) Reflectance versus normalized frequency

maximum value at a frequency slightly above $\omega = \omega_0$, and then remains high until ω reaches nearly $3\omega_0$; so the reflectance is substantial while absorption is strong. The normal dispersion region is the frequency range below ω_0 where n falls as the frequency decreases; that is, n decreases as the wavelength λ increases. The anomalous dispersion region is the frequency range above ω_0 , where n decreases as ω increases. Below ω_0 , K is small and if ϵ_{dc} is $\epsilon_r(0)$, the refractive index becomes:

$$n^2 \approx 1 + (\epsilon_{dc} - 1) \frac{\omega_0^2}{\omega_0^2 - \omega^2}; \quad \omega < \omega_0. \quad (3.7)$$

Since, $\lambda = 2\pi c/\omega$, defining $\lambda_0 = 2\pi c/\omega_0$ as the resonance wavelength, one gets:

$$n^2 \approx 1 + (\epsilon_{dc} - 1) \frac{\lambda^2}{\lambda^2 - \lambda_0^2}; \quad \lambda > \lambda_0. \quad (3.8)$$

While intuitively useful, the dispersion relation in (3.8) is far too simple. More rigorously, we have to consider the dipole oscillator quantum-mechanically, which means that a photon excites the oscillator to a higher energy level; see, for example, [3.2, 3]. The result is that we have a series of $\lambda^2/(\lambda^2 - \lambda_i^2)$ terms with various weighting factors A_i that add to unity, where the λ_i represent different resonance wavelengths. The weighting factors A_i involve quantum-mechanical matrix elements.

Figure 3.2 shows the complex relative permittivity and the complex refractive index of crystalline silicon in terms of photon energy $h\nu$. For photon energies below the bandgap energy (1.1 eV), both ϵ_r'' and K are negligible and n is close to 3.7. Both ϵ_r'' and K increase and change strongly as the photon energy becomes greater than 3 eV, far beyond the bandgap energy. Notice that both ϵ_r' and n peak at $h\nu \approx 3.5$ eV, which corresponds to a direct photoexcitation process – the excitation of electrons from the valence band to the conduction band – as discussed later.

3.1.2 Kramers–Kronig Relations

If we know the frequency dependence of the real part, ϵ_r' , of the relative permittivity of a material, then we can use the Kramers–Kronig relations to determine the frequency dependence of the imaginary part ϵ_r'' , and vice versa. The transform requires that we know the frequency dependence of either the real or imaginary part over a frequency range that is as wide as possible, ideally from dc to infinity, and that the material is linear; in

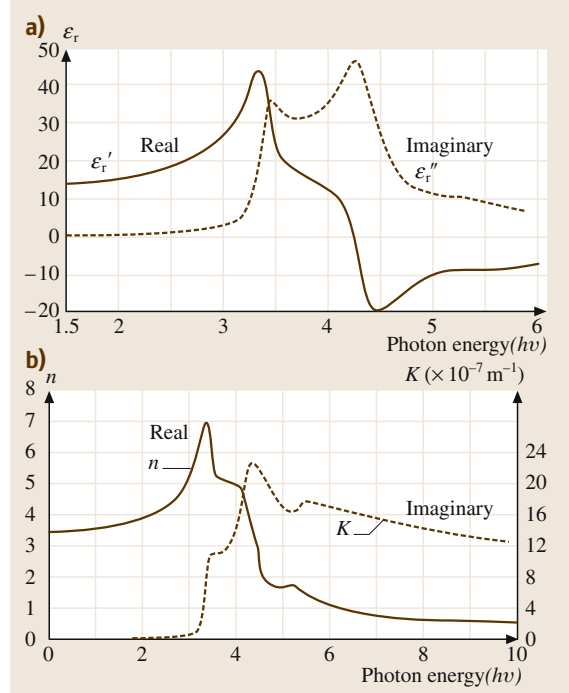


Fig. 3.2 (a) Complex relative permittivity of a silicon crystal as a function of photon energy plotted in terms of real (ϵ_r') and imaginary (ϵ_r'') parts. (b) Optical properties of a silicon crystal versus photon energy in terms of real (n) and imaginary (K) parts of the complex refractive index. After [3.4]

other words it has a relative permittivity that is independent of the applied field. The Kramers–Kronig relations for $\epsilon_r = \epsilon_r' + i\epsilon_r''$ are given by [3.4]:

$$\epsilon_r'(\omega) = 1 + \frac{2}{\pi} P \int_0^{\infty} \frac{\omega' \epsilon_r''(\omega')}{\omega'^2 - \omega^2} d\omega', \quad (3.9a)$$

and

$$\epsilon_r''(\omega) = -\frac{2\omega}{\pi} P \int_0^{\infty} \frac{\epsilon_r'(\omega')}{\omega'^2 - \omega^2} d\omega', \quad (3.9b)$$

where ω' is the integration variable, P represents the Cauchy principal value of the integral and the singularity at $\omega = \omega'$ is avoided.

Similarly, one can relate the real and imaginary parts of the polarizability, $\alpha'(\omega)$ and $\alpha''(\omega)$, and those of the

complex refractive index, $n(\omega)$ and $K(\omega)$ as well. For $\alpha'(\omega)$ and $\alpha''(\omega)$, one can analogously write:

$$\alpha'(\omega) = \frac{2}{\pi} P \int_0^{\infty} \frac{\omega' \alpha''(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (3.10a)$$

$$\alpha''(\omega') = -\frac{2\omega}{\pi} P \int_0^{\infty} \frac{\alpha'(\omega')}{\omega'^2 - \omega^2} d\omega' . \quad (3.10b)$$

3.2 Refractive Index

There are several simplified models describing the spectral dependence of the refractive index n .

3.2.1 Cauchy Dispersion Equation

The dispersion relationship for the refractive index (n) versus the wavelength of light (λ) is stated in the following form:

$$n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} , \quad (3.11)$$

where A , B and C are material-dependent specific constants. The Cauchy equation (3.11) is typically used in the visible region of the spectrum for various optical glasses, and is applied to normal dispersion. The third term is sometimes dropped for a simpler representation of n versus λ behavior. The original expression was a series in terms of the wavelength, λ , or frequency, ω , of

light as:

$$n = a_0 + a_2 \lambda^{-2} + a_4 \lambda^{-4} + a_6 \lambda^{-6} + \dots \lambda > \lambda_h , \quad (3.12a)$$

or

$$n = n_0 + n_2 \omega^2 + n_4 \omega^4 + n_6 \omega^6 + \dots \omega < \omega_h , \quad (3.12b)$$

where $\hbar\omega$ is the photon energy, $\hbar\omega_h = hc/\lambda_h$ is the optical excitation threshold (the bandgap energy), while a_0 , $a_2 \dots$, and n_0, n_2, \dots are constants. A Cauchy relation of the following form:

$$n = n_{-2}(\hbar\omega)^{-2} + n_0 + n_2(\hbar\omega)^2 + n_4(\hbar\omega)^4 , \quad (3.13)$$

can be used satisfactorily over a wide range of photon energies. The dispersion parameters, calculated from (3.13), of a few materials are listed in Table 3.1.

Table 3.1 Cauchy's dispersion parameters (obtained from (3.11)) for a few materials

Material	$\hbar\omega(\text{eV})$	$n_{-2} (\text{eV}^2)$	n_0	$n_2 (\text{eV}^{-2})$	$n_4 (\text{eV}^{-4})$
Diamond	0.05 to 5.47	-1.07×10^{-5}	2.378	8.01×10^{-3}	1.04×10^{-4}
Si	0.002 to 1.08	-2.04×10^{-8}	3.4189	8.15×10^{-2}	1.25×10^{-2}
Ge	0.002 to 0.75	-1.00×10^{-8}	4.0030	2.20×10^{-1}	1.40×10^{-1}

Table 3.2 Sellmeier coefficients of a few materials ($\lambda_1, \lambda_2, \lambda_3$ are in μm)

Material	A_1	A_2	A_3	λ_1	λ_2	λ_3
SiO ₂ (fused silica)	0.696749	0.408218	0.890815	0.0690660	0.115662	9.900559
86.5%SiO ₂ -13.5%GeO ₂	0.711040	0.451885	0.704048	0.0642700	0.129408	9.425478
GeO ₂	0.80686642	0.71815848	0.85416831	0.068972606	0.15396605	11.841931
Barium fluoride	0.63356	0.506762	3.8261	0.057789	0.109681	46.38642
Sapphire	1.023798	1.058264	5.280792	0.0614482	0.110700	17.92656
Diamond	0.3306	4.3356		0.175	0.106	
Quartz, n_0	1.35400	0.010	0.9994	0.092612	10.700	9.8500
Quartz, n_e	1.38100	0.0100	0.9992	0.093505	11.310	9.5280
KDP, n_0	1.2540	0.0100	0.0992	0.09646	6.9777	5.9848
KDP, n_e	1.13000	0.0001	0.9999	0.09351	7.6710	12.170

Cauchy's dispersion relation, given in (3.13), was originally called the elastic-ether theory of the refractive index [3.5–7]. It has been widely used for many materials, although in recent years it has been largely replaced by the Sellmeier equation, which we consider next.

3.2.2 Sellmeier Dispersion Equation

The dispersion relationship can be quite complicated in practice. An example of this is the Sellmeier equation, which is an empirical relation between the refractive index n of a substance and the wavelength λ of light in the form of a series of $\lambda^2/(\lambda^2 - \lambda_i^2)$ terms, given by:

$$n^2 = 1 + \frac{A_1\lambda^2}{\lambda^2 - \lambda_1^2} + \frac{A_2\lambda^2}{\lambda^2 - \lambda_2^2} + \frac{A_3\lambda^2}{\lambda^2 - \lambda_3^2} + \dots, \quad (3.14)$$

where λ_i is a constant and $A_1, A_2, A_3, \lambda_1, \lambda_2$ and λ_3 are called Sellmeier coefficients, which are determined by fitting this expression to the experimental data. The full Sellmeier formula has more terms of similar form, such as $A_i\lambda^2/(\lambda^2 - \lambda_i^2)$, where $i = 4, 5, \dots$ but these can generally be neglected when considering n versus λ behavior over typical wavelengths of interest and by ensuring that the three terms included in the Sellmeier equation correspond to the most important or relevant terms in the summation. Examples of Sellmeier coefficients for some materials, including pure silica (SiO_2) and 86.5 mol.% SiO_2 -13.5 mol.% GeO_2 , are given in Table 3.2. Two methods are used to find the refractive index of silica-germania glass (SiO_2)_{1-x}(GeO_2)_x: (a) a simple, but approximate, linear interpolation of the refractive index between known compositions, for example $n(x) - n(0.135) = (x - 0.135)[n(0.135) - n(0)]/0.135$ for (SiO_2)_{1-x}(GeO_2)_x, so $n(0.135)$ is used for 86.5 mol.% SiO_2 -13.5 mol.% GeO_2 and $n(0)$ is used for SiO_2 ; (b) an interpolation for the coefficients A_i and λ_i between SiO_2 and GeO_2 :

$$n^2 - 1 = \frac{\{A_1(S) + X[A_1(G) - A_1(S)]\}\lambda^2}{\lambda^2 - \{\lambda_1(S) + X[\lambda_1(G) - \lambda_1(S)]\}^2} + \dots, \quad (3.15)$$

where X is the atomic fraction of germania, S and G in parentheses refer to silica and germania [3.10]. The theoretical basis of the Sellmeier equation is that the solid is represented as a sum of N lossless (frictionless) Lorentz oscillators such that each takes the form of $\lambda^2/(\lambda^2 - \lambda_i^2)$ with different λ_i , and each has different strengths, with weighting factors ($A_i, i = 1$ to N) [3.11, 12]. Knowledge of appropriate dispersion relationships is essential when designing photonic devices, such as waveguides.

There are other dispersion relationships that inherently take account of various contributions to optical properties, such as the electronic and ionic polarization and the interactions of photons with free electrons. For example, for many semiconductors and ionic crystals, two useful dispersion relations are:

$$n^2 = A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E}, \quad (3.16)$$

and

$$n^2 = A + \frac{B}{\lambda^2 - \lambda_0^2} + \frac{C}{(\lambda^2 - \lambda_0^2)^2} + D\lambda^2 + E\lambda^4, \quad (3.17)$$

where A, B, C, D, E and λ_0 are constants particular to a given material. Table 3.3 provides a few examples.

The refractive index of a semiconductor material typically decreases with increasing bandgap energy E_g . There are various empirical and semi-empirical rules and expressions that relate n to E_g . In *Moss' rule*, n and E_g are related by $n^4 E_g = K = \text{constant}$ (≈ 100 eV). In the *Hervé-Vandamme* relationship [3.13],

$$n^2 = 1 + \left(\frac{A}{E_g + B} \right)^2, \quad (3.18)$$

where A and B are constants ($A \approx 13.6$ eV and $B \approx 3.4$ eV and $\text{dB}/\text{dT} \approx 2.5 \times 10^{-5}$ eV/K). The refractive index typically increases with increasing temperature. The temperature coefficient of the refractive index (TCRI) of a semiconductor can be found from the Hervé-Vandamme relationship [3.13]:

$$\text{TCRI} = \frac{1}{n} \cdot \frac{dn}{dT} = \frac{(n^2 - 1)^{3/2}}{13.6n^2} \left(\frac{dE_g}{dT} + \frac{dB}{dT} \right). \quad (3.19)$$

Table 3.3 Parameters from Eqs. (3.16) and (3.17) for some selected materials (Si data from [3.8]; others from [3.9])

Material	$\lambda_0(\mu\text{m})$	A	$B(\mu\text{m})^2$	$C(\mu\text{m})^{-4}$	$D(\mu\text{m})^{-2}$	$E(\mu\text{m})^{-4}$
Silicon	0.167	3.41983	0.159906	-0.123109	1.269×10^{-6}	-1.951×10^{-9}
MgO	0.11951	2.95636	0.021958	0	-1.0624×10^{-2}	-2.05×10^{-5}
LiF	0.16733	1.38761	0.001796	-4.1×10^{-3}	-2.3045×10^{-3}	-5.57×10^{-6}
AgCl	0.21413	4.00804	0.079009	0	-8.5111×10^{-4}	-1.976×10^{-7}

Table 3.4 Examples of parameters for the Wemple–DiDomenico dispersion relationship (3.23), for various materials

Material	N_c	Z_a	N_c	$E_0(\text{eV})$	$E_d(\text{eV})$	$\beta(\text{eV})$	β	Comment
NaCl	6	1	8	10.3	13.6	0.28	β_i	Halides, LiF, NaF, etc.
CsCl	8	1	8	10.6	17.1	0.27	β_i	CsBr, CsI, etc.
TiCl	8	1	10	5.8	20.6	0.26	β_i	TlBr
CaF ₂	8	1	8	15.7	15.9	0.25	β_i	BaF ₂ , etc.
CaO	6	2	8	9.9	22.6	0.24	β_i	Oxides, MgO, TeO ₂ , etc.
Al ₂ O ₃	6	2	8	13.4	27.5	0.29	β_i	
LiNbO ₃	6	2	8	6.65	25.9	0.27	β_i	
TiO ₂	6	2	8	5.24	25.7	0.27	β_i	
ZnO	4	2	8	6.4	17.1	0.27	β_i	
ZnSe	4	2	8	5.54	27	0.42	β_c	II–VI, Zinc blende, ZnS, ZnTe, CdTe
GaAs	4	3	8	3.55	33.5	0.35	β_c	III–V, Zinc blende, GaP, etc.
Si (crystal)	4	4	8	4.0	44.4	0.35	β_c	Diamond, covalent bonding; C (diamond), Ge, β -SiC etc.
SiO ₂ (crystal)	4	2	8	13.33	18.10	0.28	β_i	Average crystalline form
SiO ₂ (amorphous)	4	2	8	13.38	14.71	0.23	β_i	Fused silica
CdSe	4	2	8	4.0	20.6	0.32	$\beta_i - \beta_c$	Wurtzite

TCRI is typically in the range 10^{-6} to 10^{-4} .

3.2.3 Gladstone–Dale Formula

The *Gladstone–Dale formula* is an empirical equation that allows the average refractive index n of an oxide glass to be calculated from its density ρ and its constituents as:

$$\frac{n-1}{\rho} = p_1 k_1 + p_2 k_2 + \dots = \sum_{i=1}^N p_i k_i = C_{GD}, \quad (3.20)$$

where the summation is for various oxide components (each a simple oxide), p_i is the weight fraction of the i -th oxide in the compound, and k_i is the refraction coefficient that represents the polarizability of the i -th oxide. The right hand side of (3.20) is called the Gladstone–Dale coefficient C_{GD} . In more general terms, as a mixture rule for the overall refractive index, the Gladstone–Dale formula is frequently written as:

$$\frac{n-1}{\rho} = \frac{n_1-1}{\rho_1} w_1 + \frac{n_2-1}{\rho_2} w_2 + \dots, \quad (3.21)$$

where n and ρ are the effective refractive index and effective density of the whole mixture, n_1 , n_2 ,

... are the refractive indices of the constituents, and ρ_1 , ρ_2 , ... represent the densities of each constituent. Gladstone–Dale equations for the polymorphs of SiO₂ and TiO₂ give the average n values as [3.14]:

$$n(\text{SiO}_2) = 1 + 0.21\rho \quad \text{and} \quad n(\text{TiO}_2) = 1 + 0.40\rho. \quad (3.22)$$

3.2.4 Wemple–Di Dominico Dispersion Relation

The *Wemple–Di Dominico dispersion relation* is a semi-empirical single oscillator-based relationship used to find the refractive indices of a variety of materials for photon energies below the interband absorption edge, given by

$$n^2 = 1 + \frac{E_0 E_d}{E_0^2 - (h\nu)^2}, \quad (3.23)$$

where ν is the frequency, h is the Planck constant, E_0 is the single oscillator energy, E_d is the dispersion energy, which is a measure of the average strength of interband optical transitions; $E_d = \beta N_c Z_a N_c$ (eV), where N_c is the effective coordination number of the cation nearest-neighbor to the anion ($N_c = 6$ in NaCl, $N_c = 4$

in Ge), Z_a is the formal chemical valency of the anion ($Z_a = 1$ in NaCl; 2 in Te; 3 in GaP), N_e is the effective number of valence electrons per anion excluding the cores ($N_e = 8$ in NaCl, Ge; 10 in TlCl; 12 in Te; $9\frac{1}{3}$ in As_2Se_3), and β is a constant that depends on whether the interatomic bond is ionic (β_i) or covalent (β_c): $\beta_i = 0.26 \pm 0.04$ eV (this applies to halides such as NaCl and ThBr and most oxides, including Al_2O_3 , for example), while $\beta_c = 0.37 \pm 0.05$ eV (applies to tetrahedrally bonded $A^N B^{8-N}$ zinc blende- and diamond-type structures such as GaP, ZnS, for instance). (Note that wurtzite crystals have a β that is intermediate between β_i and β_c .) Also, empirically, $E_0 = CE_g(D)$, where $E_g(D)$ is the lowest direct bandgap and C is a constant, typically ≈ 1.5 . E_0 has been associated with the main peak in the $\epsilon''_r(h\nu)$ versus $h\nu$ spectrum. The parameters required to calculate n from (3.23) are listed in Table 3.4 [3.15].

3.2.5 Group Index (N_g)

The group index represents the factor by which the group velocity of a group of waves in a dielectric medium is reduced with respect to propagation in free space,

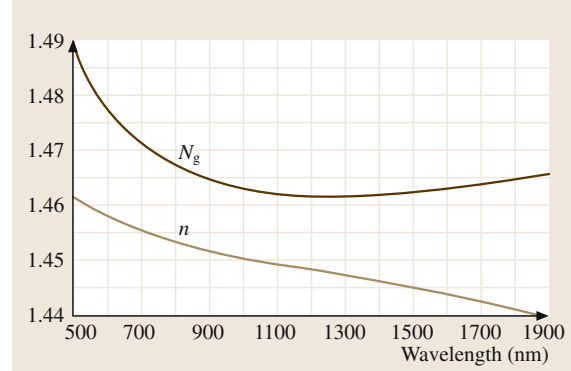


Fig. 3.3 Refractive index n and the group index N_g of pure SiO_2 (silica) glass as a function of wavelength

$N_g = c/v_g$, where v_g is the group velocity. The group index can be determined from the ordinary refractive index n via

$$N_g = n - \lambda \frac{dn}{d\lambda}, \quad (3.24)$$

where λ is the wavelength of light. The relation between N_g and n for SiO_2 is illustrated in Fig. 3.3.

3.3 Optical Absorption

The main optical properties of a semiconductor are typically its refractive index n and its extinction coefficient K or absorption coefficient α (or equivalently the real and imaginary parts of the relative permittivity), as well

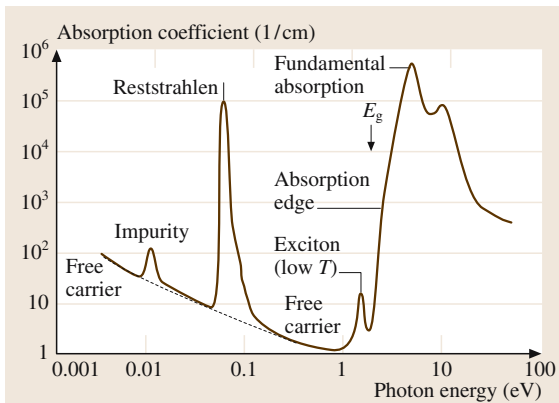


Fig. 3.4 Absorption coefficient plotted as a function of the photon energy in a typical semiconductor, illustrating various possible absorption processes

as their dispersion relations (their dependence on the electromagnetic radiation wavelength λ or photon energy $h\nu$) and the changes in the dispersion relations with temperature, pressure, alloying, impurities, and so on. A typical relationship between the absorption coefficient and the photon energy observed in a crystalline semiconductor is shown in Fig. 3.4, where various possible absorption processes are illustrated. The important features of the behavior of the α versus $h\nu$ can be summarized as follows. (a) Free carrier absorption due to the presence of free electrons and holes, an effect that decreases with increasing photon energy. (b) An impurity absorption band (usually narrow) due to the various dopants. (c) Reststrahlen or lattice absorption in which the radiation is absorbed by vibrations of the crystal ions. (d) Exciton absorption peaks that are usually observed at low temperatures and are close to the fundamental absorption edge. (e) Band-to-band or fundamental absorption of photons, which excites an electron from the valence to the conduction band. Type (e) absorption has a large absorption coefficient and occurs when the photon energy reaches the bandgap energy

Table 3.5 Crystal structure, lattice parameter a , bandgap energy E_g at 300 K, type of bandgap (D=direct and I=indirect), change in E_g per unit temperature change (dE_g/dT) at 300 K, bandgap wavelength λ_g and refractive index n close to λ_g (A=amorphous, D=diamond, W=wurtzite, ZB=zinc blende). Approximate data from various sources

Semiconductors	Crystal	a (nm)	E_g (eV)	Type	dE_g/dT (meV/K)	λ_g (μm)	$n(\lambda_g)$	dn/dT (10^{-5} K^{-1})
Group IV								
Diamond	D	0.3567	5.48	I	−0.05	0.226	2.74	1.1
Ge	D	0.5658	0.66	I	−0.37	1.87	4	27.6 42.4 (4 μm)
Si	D	0.5431	1.12	I	−0.25	1.11	3.45	13.8 16 (5 μm)
a-Si:H	A		1.7–1.8			0.73		
SiC(α)	W	0.3081 a 1.5120 c	2.9	I	−0.33	0.42	2.7	9
III–V Compounds								
AlAs	ZB	0.5661	2.16	I	−0.50	0.57	3.2	15
AlP	ZB	0.5451	2.45	I	−0.35	0.52	3	11
AlSb	ZB	0.6135	1.58	I	−0.3	0.75	3.7	
GaAs	ZB	0.5653	1.42	D	−0.45	0.87	3.6	15
GaAs _{0.88} Sb _{0.12}	ZB		1.15	D		1.08		
GaN	W	0.3190 a 0.5190 c	3.44	D	−0.45	0.36	2.6	6.8
GaP	ZB	0.5451	2.26	I	−0.54	0.40	3.4	
GaSb	ZB	0.6096	0.73	D	−0.35	1.7	4	33
In _{0.53} Ga _{0.47} As on InP	ZB	0.5869	0.75	D		1.65		
In _{0.58} Ga _{0.42} As _{0.9} P _{0.1} on InP	ZB	0.5870	0.80	D		1.55		
In _{0.72} Ga _{0.28} As _{0.62} P _{0.38} on InP	ZB	0.5870	0.95	D		1.3		
InP	ZB	0.5869	1.35	D	−0.40	0.91	3.4–3.5	9.5
InAs	ZB	0.6058	0.36	D	−0.28	3.5	3.8	2.7
InSb	ZB	0.6479	0.18	D	−0.3	7	4.2	29
II–VI Compounds								
ZnSe	ZB	0.5668	2.7	D	−0.50	0.46	2.3	6.3
ZnTe	ZB	0.6101	2.3	D	−0.45	0.55	2.7	

E_g . It is probably the most important absorption effect; its characteristics for $h\nu > E_g$ can be predicted using the results of Sect. 3.3.3. The values of E_g , and its temperature shift, dE_g/dT , are therefore important factors in semiconductor-based optoelectronic devices. In nearly all semiconductors E_g decreases with temperature, hence shifting the fundamental absorption to longer wavelengths. The refractive index n also changes with temperature. dn/dT depends on the wavelength, but for many semiconductors $(dn/dT)/n \approx 5 \times 10^{-5} \text{ K}^{-1}$; for example, for GaAs, $(dn/dT)/n = 4 \times 10^{-4} \text{ K}^{-1}$ at

$\lambda = 2 \mu\text{m}$. There is a good correlation between the refractive indices and the bandgaps of semiconductors in which, typically, n decreases as E_g increases; semiconductors with wider bandgaps have lower refractive indices. The refractive index n and the extinction coefficient K (or α) are related by the Kramers–Kronig relations. Thus, large increases in the absorption coefficient for $h\nu$ near and above the bandgap energy E_g also result in increases in the refractive index n versus $h\nu$ in this region. Optical (and some structural) properties of various semiconductors are listed in Table 3.5.

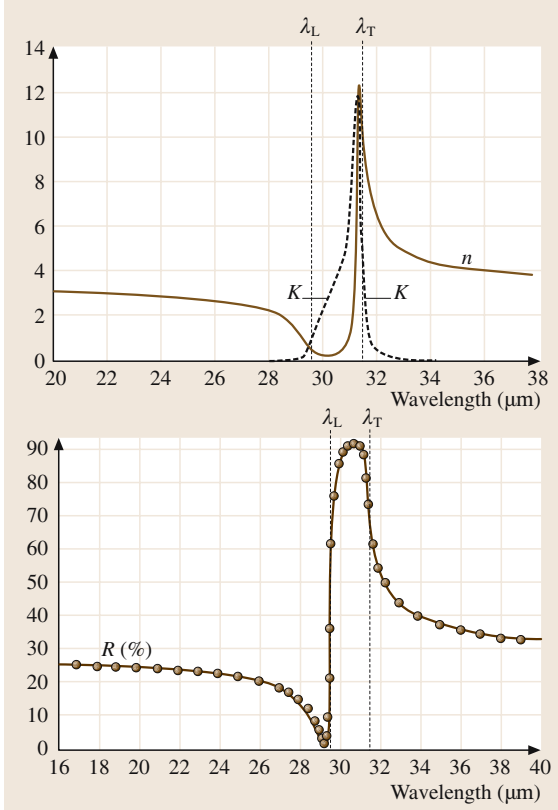


Fig. 3.5 Infrared refractive index n , extinction coefficient K (top), and reflectance R (bottom) of AlSb. Note that the wavelength axes are not identical, and the wavelengths λ_T and λ_L , corresponding to ω_T and ω_L , respectively, are shown as dashed vertical lines. After [3.16]

3.3.1 Lattice or Reststrahlen Absorption and Infrared Reflection

In the infrared wavelength region, ionic crystals reflect and absorb light strongly due to the resonance interaction of the electromagnetic (EM) wave field with the transverse optical (TO) phonons. The dipole oscillator model based on ions driven by an EM wave results in

$$\varepsilon_r = \varepsilon'_r - i\varepsilon''_r = \varepsilon_{r\infty} + \frac{\varepsilon_{r0} - \varepsilon_{r\infty}}{\left(\frac{\omega}{\omega_T}\right)^2 - 1 + i\frac{\gamma}{\omega_T}\left(\frac{\omega}{\omega_T}\right)}, \quad (3.25)$$

where ε_{r0} and $\varepsilon_{r\infty}$ are the relative permittivity at $\omega = 0$ (very low frequencies) and $\omega = \infty$ (very high frequencies) respectively, γ is the loss coefficient per unit reduced mass, representing the rate of energy transfer

from the EM wave to optical phonons, and ω_0 is a resonance frequency that is related to the “spring” constant between the ions. By definition, the frequency ω_T is

$$\omega_T^2 = \omega_0^2 \left(\frac{\varepsilon_{r\infty} + 2}{\varepsilon_{r0} + 2} \right). \quad (3.26)$$

The loss ε''_r and the absorption are maxima when $\omega = \omega_T$, and the wave is attenuated by the transfer of energy to the transverse optical phonons, so the EM wave couples to the transverse optical phonons. At $\omega = \omega_L$, the wave couples to the longitudinal optical (LO) phonons. Figure 3.5 shows the optical properties of AlSb [3.16] in terms of n , K and R versus wavelength. The peaks in the extinction coefficient K and reflectance R occur over roughly the same wavelength region, corresponding to the coupling of the EM wave to the transverse optical phonons. At wavelengths close to $\lambda_T = 2\pi/\omega_T$, n and K peak, and there is strong absorption of light (which corresponds to the EM wave resonating with the TO lattice vibrations), and then R rises sharply.

3.3.2 Free Carrier Absorption (FCA)

An electromagnetic wave with sufficiently low frequency oscillations can interact with free carriers in a material and thereby drift the carriers. This interaction results in an energy loss from the EM wave to the lattice vibrations through the carrier scattering processes. Based on the Drude model, the relative permittivity $\varepsilon_r(\omega)$ due to N free electrons per unit volume is given by

$$\varepsilon_r = \varepsilon'_r - i\varepsilon''_r = 1 - \frac{\omega_p^2}{\omega^2 - i\omega/\tau};$$

$$\omega_p^2 = \frac{Ne^2}{\varepsilon_0 m_e}, \quad (3.27)$$

where ω_p is a plasma frequency which depends on the electron concentration, while τ is the relaxation time of the scattering process (the mean scattering time). For metals where the electron concentration is very large, ω_p is of the order of $\approx 10^{16}$ rad/s, at UV frequencies, and for $\omega > \omega_p$, $\varepsilon_r \approx 1$, and the reflectance becomes very small. Metals lose their free-electron reflectance in the UV range, thus becoming UV-transparent. The reflectance does not fall to zero because there are other absorption processes such as interband electron excitations or excitations from core levels to energy bands. Plasma edge transparency, where the reflectance almost vanishes, can also be observed in doped semiconductors. For example, the reflectance of doped InSb has a plasma edge wavelength that decreases with increasing free carrier concentration [3.17]. Equation (3.27) can be written

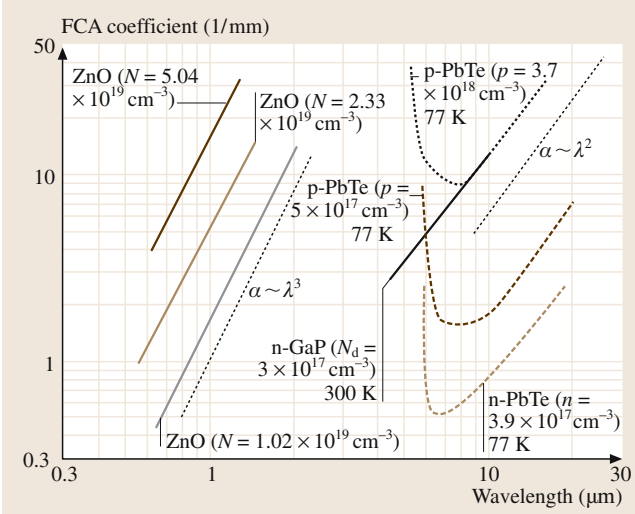


Fig. 3.6 Free carrier absorption in n-GaP at 300 K [3.18], p- and n-type PbTe [3.19] at 77 K and In doped n-type ZnO at room temperature. After [3.20]

in terms of the conductivity σ_0 at low frequencies (dc) as

$$\epsilon_r = \epsilon'_r - i\epsilon''_r = 1 - \frac{\tau\sigma_0}{\epsilon_0[(\omega\tau)^2 + 1]} - i\frac{\sigma_0}{\epsilon_0\omega[(\omega\tau)^2 + 1]}. \quad (3.28)$$

In metals, σ_0 is high. At frequencies where $\omega < 1/\tau$, the imaginary part $\epsilon''_r = \sigma_0/\epsilon_0\omega$ is normally much more than 1, and $n = K \approx \sqrt{(\epsilon''_r/2)}$, so that the free carrier attenuation coefficient α is then given by

$$\alpha = 2k_0K \approx \frac{2\omega}{c} \left(\frac{\epsilon''_r}{2} \right)^{1/2} \approx (2\sigma_0\mu_0)^{1/2} \omega^{1/2}. \quad (3.29)$$

Furthermore, the reflectance can also be calculated using $n = K \approx \sqrt{(\epsilon''_r/2)}$, which leads to the well-known Hagen–Rubens relationship [3.21]

$$R \approx 1 - 2 \left(\frac{2\omega\epsilon_0}{\sigma_0} \right)^{1/2}. \quad (3.30)$$

In semiconductors, one typically encounters $\sigma_0/\epsilon_0\omega < 1$, since the free electron concentration is small, and we can treat n as constant due to various other polarization mechanisms such as electronic polarization. Since $2nK = \epsilon''_r$, the absorption coefficient becomes [3.22]:

$$\alpha = 2k_0K \approx \frac{2\omega}{c} \left(\frac{\epsilon''_r}{2n} \right) = \frac{\sigma_0}{nc\epsilon_0[(\omega\tau)^2 + 1]}. \quad (3.31)$$

At low frequencies where $\omega < 1/\tau$, we have $\alpha(\lambda) \propto \sigma_0/n(\lambda)$ so that α should be controlled by the dc conductivity, and hence the amount of doping. Furthermore, α will exhibit the frequency dependence of the refractive index n ($\alpha(\lambda) \propto 1/n(\lambda)$), in which n will typically be determined by the electronic polarization of the crystal.

At high frequencies where $\omega > 1/\tau$,

$$\alpha \propto \sigma_0/\omega^2 \propto N\lambda^2, \quad (3.32)$$

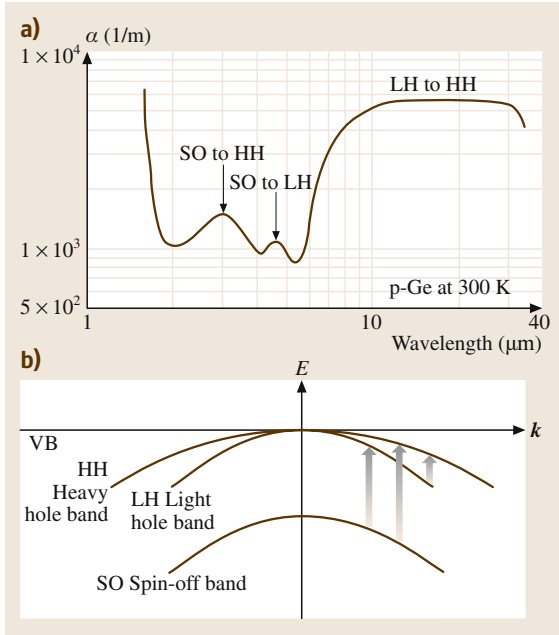
where α is proportional to N , the free carrier concentration, and λ^2 . Experimental observations on FCA in doped semiconductors are in general agreement with these predictions. For example, α increases with N , whether N is increased by doping or by carrier injection [3.23]. However, not all semiconductors show the simple $\alpha \propto \lambda^2$ behavior. A proper account of the field-driven electron motion and scattering must consider the fact that τ will depend on the electron energy. The correct approach is to use the Boltzmann transport equation [3.24] with the appropriate scattering mechanism. FCA can be calculated using a quantum-mechanical approach based on second-order time-dependent perturbation theory with Fermi–Dirac statistics [3.25].

Absorption due to free carriers is commonly written as $\alpha \propto \lambda^p$, where the index p depends primarily on the scattering mechanism, although it is also influenced by the compensation doping ratio, if the semiconductor has been doped by compensation, and the free carrier concentration. In the case of lattice scattering, one must consider scattering from acoustic and optical phonons. For acoustic phonon scattering $p \approx 1.5$, for optical phonon scattering $p \approx 2.5$, and for impurity scattering $p \approx 3.5$. The observed free carrier absorption coefficient will then have all three contributions:

$$\alpha = A_{\text{acoustic}}\lambda^{1.5} + A_{\text{optical}}\lambda^{2.5} + A_{\text{impurity}}\lambda^{3.5}. \quad (3.33)$$

Inasmuch as α for FCA depends on the free carrier concentration N , it is possible to evaluate the latter from the experimentally measured α , given its wavelength dependence and p as discussed by [3.26]. The free carrier absorption coefficients $\alpha(\text{mm}^{-1})$ for GaP, n-type PbTe and n-type ZnO are shown in Fig. 3.6.

Free carrier absorption in p-type Ge demonstrates how the FCA coefficient α can be dramatically different than what is expected from (3.33). Figure 3.7a shows the wavelength dependence of the absorption coefficient for p-Ge over the wavelength range from about 2 to 30 μm [3.27]. The observed absorption is due to excitations of electrons from the spin-off band to the heavy



hole band, from the spin-off band to the light hole band, and from the light hole band to the heavy hole band, as marked in the Fig. 3.7b.

3.3.3 Band-to-Band or Fundamental Absorption

Crystalline Solids

Band-to-band absorption or fundamental absorption of radiation is due to the photoexcitation of an electron from the valence band to the conduction band. There are

Fig. 3.7 (a) Free carrier absorption due to holes in p-Ge [3.27]. (b) The valence band of Ge has three bands; heavy hole, light hole and a spin-off bands

two types of band-to-band absorptions, corresponding to direct and indirect transitions.

A direct transition is a photoexcitation process in which no phonons are involved. The photon momentum is negligible compared with the electron momentum, so that when the photon is absorbed, exciting an electron from the valence band (VB) to the conduction band (CB), the electron's k -vector does not change. A direct transition on an $E-k$ diagram is a vertical transition from an initial energy E and wavevector k in the VB to a final energy E' and a wavevector k' in the CB where $k' = k$, as shown in Fig. 3.8a. The energy $(E' - E_c)$ is the kinetic energy $(\hbar k)^2/(2m_e^*)$ of the electron with an effective mass m_e^* , and $(E_v - E)$ is the kinetic energy $(\hbar k)^2/(2m_h^*)$ of the hole left behind in the VB. The ratio of the kinetic energies of the photogenerated electron and hole depends inversely on the ratio of their effective masses.

The absorption coefficient α is derived from the quantum-mechanical probability of transition from E to E' , the occupied density of states at E in the VB from which electrons are excited, and the unoccupied density of states in the CB at $E + \hbar\nu$. Thus, α depends on the joint density of states at E and $E + \hbar\nu$, and we have to suitably integrate this joint density of states. Near the band edges, the density of states can be approximated by a parabolic band, and α rises with the photon energy following

$$\alpha \hbar\nu = A(\hbar\nu - E_g)^{1/2}, \quad (3.34)$$

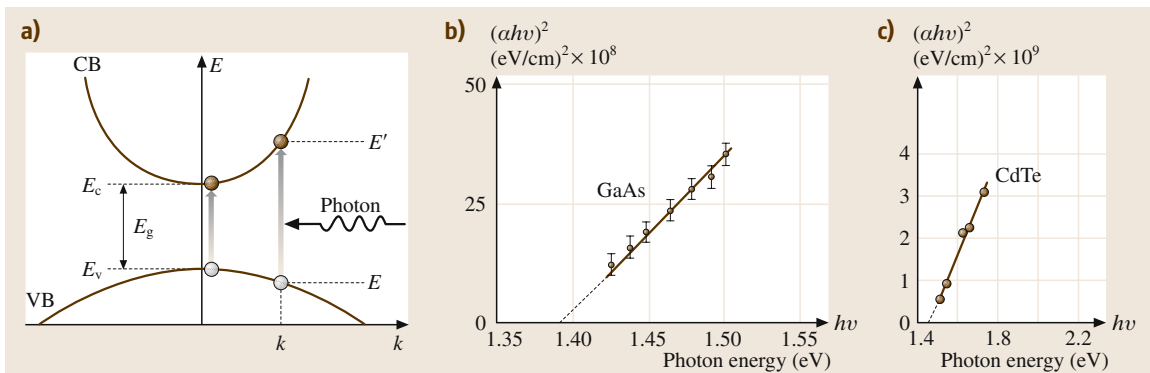


Fig. 3.8 (a) A direct transition from the valence band (VB) to the conduction band (CB) through the absorption of a photon. Absorption behavior represented as $(\alpha \hbar\nu)^2$ versus photon energy $\hbar\nu$ near the band edge for single crystals of (b) p-type GaAs, from [3.28] and (c) CdTe. After [3.29]

where the constant $A \approx [e^2/(nch^2m_e^*)](2\mu^*)^{3/2}$ in which μ^* is a reduced electron and hole effective mass, n is the refractive index, and E_g is the direct bandgap, with minimum $E_c - E_v$ at the same k value. Experiments indeed show this type of behavior for photon energies above E_g and close to E_g , as shown in Fig. 3.8b for a GaAs crystal [3.28] and in (c) for a CdTe crystal [3.29]. The extrapolation to zero photon energy gives the direct bandgap E_g , which is about 1.40 eV for GaAs and 1.46–1.49 eV for CdTe. For photon energies very close to the bandgap energy, the absorption is usually due to exciton absorption, especially at low temperatures, and is discussed later in this chapter.

In indirect bandgap semiconductors, such as Si and Ge, the photon absorption for photon energies near E_g requires the absorption and emission of phonons during the absorption process, as illustrated in Fig. 3.9a. The absorption onset corresponds to a photon energy of $(E_g - h\vartheta)$, which represents the absorption of a phonon with energy $h\vartheta$. For the latter, α is proportional to $[h\nu - (E_g - h\vartheta)]^2$. Once the photon energy reaches $(E_g + h\vartheta)$, then the photon absorption process can also occur by phonon emission, for which the absorption coefficient is larger than that for phonon absorption. The absorption coefficients for the phonon absorption and emission processes are given by [3.30]

$$\alpha_{\text{absorption}} = A[f_{\text{BE}}(h\vartheta)][h\nu - (E_g - h\vartheta)]^2; \quad h\nu > (E_g - h\vartheta), \quad (3.35)$$

and

$$\alpha_{\text{emission}} = A[1 - f_{\text{BE}}(h\vartheta)][h\nu - (E_g + h\vartheta)]^2; \quad h\nu > (E_g + h\vartheta), \quad (3.36)$$

where A is a constant and $f_{\text{BE}}(h\vartheta)$ is the Bose–Einstein distribution function at the phonon energy $h\vartheta$, $f_{\text{BE}}(h\vartheta) = [\exp(h\vartheta/k_B T) - 1]^{-1}$, where k_B is the Boltzmann constant and T is the temperature. As we increase the photon energy in the range $(E_g - h\vartheta) < h\nu < (E_g + h\vartheta)$, the absorption is controlled by $\alpha_{\text{absorption}}$ and the plot of $\alpha^{1/2}$ versus $h\nu$ has an intercept of $(E_g - h\vartheta)$.

On the other hand, for $h\nu > (E_g + h\vartheta)$, the overall absorption coefficient is $\alpha_{\text{absorption}} + \alpha_{\text{emission}}$, but at slightly higher photon energies than $(E_g + h\vartheta)$, α_{emission} quickly dominates over $\alpha_{\text{absorption}}$ since $[f_{\text{BE}}(h\vartheta)] \gg [1 - f_{\text{BE}}(h\vartheta)]$. Figure 3.9b shows the behavior of $\alpha^{1/2}$ versus photon energy for Si at two temperatures for $h\nu$ near band edge absorption. At low temperatures, $f_{\text{BE}}(h\vartheta)$ is small and $\alpha_{\text{absorption}}$ decreases with decreasing temperature, as apparent from Fig. 3.9b.

Equations (3.35) and (3.36) intersect the photon energy axis at $(E_g - h\vartheta)$ and $(E_g + h\vartheta)$, which can be used to obtain E_g .

An examination of the extinction coefficient K or ϵ_r'' versus photon energy for Si in Fig. 3.2 shows that absorption peaks at certain photon energies, $h\nu \approx 3.5$ and 4.3 eV. These peaks are due to the fact that the joint density of states function peaks at these energies. The absorption coefficient peaks whenever there is a direct transition in which the E versus k curve in the VB is parallel to the E versus k curve in the CB, as schematically illustrated in Fig. 3.9a, where a photon of energy $h\nu_{12}$ excites an electron from state 1 in the VB to state 2 in the CB in a direct transition, i.e. $k_1 = k_2$. Such transitions where E versus k curves are parallel at a photon energy $h\nu_{12}$ result in a peak in the absorption versus photon energy behavior, and can be represented by the condition that

$$(\nabla_k E)_{\text{CB}} - (\nabla_k E)_{\text{VB}} = 0. \quad (3.37)$$

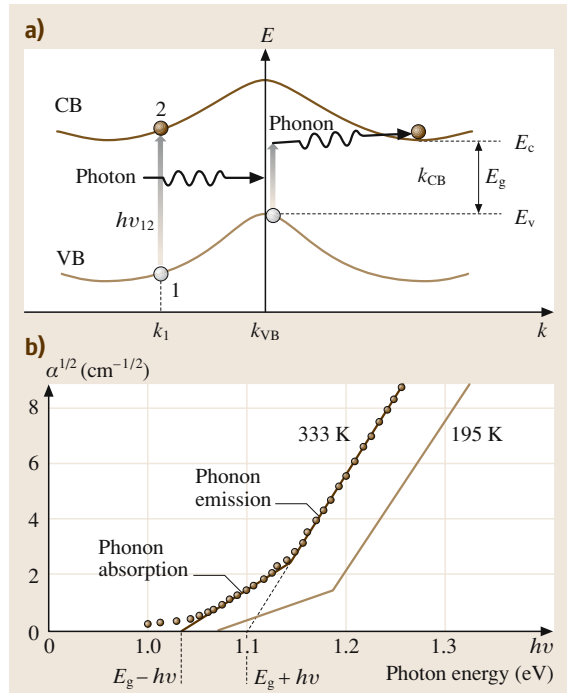


Fig. 3.9 (a) Indirect transitions across the bandgap involve phonons. Direct transitions in which dE/dk in the CB is parallel to dE/dk in the VB lead to peaks in the absorption coefficient. (b) Fundamental absorption in Si at two temperatures. The overall behavior is well described by (3.35) and (3.36)

The above condition is normally interpreted as the joint density of states reaching a peak value at certain points in the Brillouin zone called van Hove singularities. Identification of peaks in K versus $h\nu$ involves the examination of all E versus k curves of a given crystal that can participate in a direct transition. The silicon ε_r'' peaks at $h\nu \approx 3.5$ eV and 4.3 eV correspond to (3.37) being satisfied at points L, along $\langle 111 \rangle$ in k -space, and X along $\langle 100 \rangle$ in k -space, at the edges of the Brillouin zone.

In degenerate semiconductors, the Fermi level E_F is in a band; for example, in the CB for a degenerate n -type semiconductor. Electrons in the VB can only be excited to states above E_F in the CB rather than to the bottom of the CB. The absorption coefficient then depends on the free carrier concentration since the latter determines E_F . Fundamental absorption is then said to depend on band filling, and there is an apparent shift in the absorption edge, called the *Burstein–Moss shift*. Furthermore, in degenerate indirect semiconductors, the indirect transition may involve a non-phonon scattering process, such as impurity or electron–electron scattering, which can change the electron’s wavevector k . Thus, in degenerate indirect bandgap semiconductors, absorption can occur without phonon assistance and the absorption coefficient becomes:

$$\alpha \propto [h\nu - (E_g + \Delta E_F)]^2, \quad (3.38)$$

where ΔE_F is the energy depth of E_F into the band measured from the band edge.

Heavy doping of degenerate semiconductors normally leads to a phenomenon called bandgap narrowing and bandtailing. Bandtailing means that the band edges at E_v and E_c are no longer well defined cut-off energies, and there are electronic states above E_v and below E_c where the density of states falls sharply with energy away from the band edges. Consider a degenerate direct band gap p -type semiconductor. One can excite electrons from states below E_F in the VB, where the band is nearly parabolic, to tail states below E_c , where the density of states decreases exponentially with energy into the bandgap, away from E_c . Such excitations lead to α depending exponentially on $h\nu$, a dependence that is called the *Urbach rule* [3.31, 32], given by:

$$\alpha = \alpha_0 \exp[(h\nu - E_0)/\Delta E] \quad (3.39)$$

where α_0 and E_0 are material-dependent constants, and ΔE , called the Urbach width, is also a material-dependent constant. The Urbach rule was originally reported for alkali halides. It has been observed for many

ionic crystals, degenerately doped crystalline semiconductors, and almost all amorphous semiconductors. While exponential bandtailing can explain the observed Urbach tail of the absorption coefficient versus photon energy, it is also possible to attribute the absorption tail behavior to strong internal fields arising, for example, from ionized dopants or defects.

Amorphous Solids

In a defect-free crystalline semiconductor, a well-defined energy gap exists between the valence and conduction bands. In contrast, in an amorphous semiconductor, the distributions of conduction band and valence band electronic states do not terminate abruptly at the band edges. Instead, some electronic states called the tail states encroach into the gap region [3.33]. In addition to tail states, there are also other localized states deep within the gap region [3.34]. These localized tail states in amorphous semiconductors are contributed by defects. The defects in amorphous semiconductors are considered to be all cases of departure from the normal nearest-neighbor coordination (or normal valence requirement). Examples of defects are: broken and dangling bonds (typical for amorphous silicon), over- and under-coordinated atoms (such as “valence alternation pairs” in chalcogenide glasses), voids, pores, cracks and other macroscopic defects. Mobility edges exist, which separate these localized states from their extended counterparts; tail and deep defect states are localized [3.35–37]. These localized tail and deep defect states are responsible for many of the unique properties exhibited by amorphous semiconductors.

Despite years of intensive investigation, the exact form of the distribution of electronic states associated with amorphous semiconductors remains a subject of some debate. While there are still some unresolved theoretical issues, there is general consensus that the tail states arise as a consequence of the disorder present within amorphous networks, and that the width of these tails reflects the amount of disorder present [3.38]. Experimental results (from, for example, [3.39, 40]) suggest exponential distributions for the valence and conduction band tail states in a-Si:H, although other possible functional forms [3.41] cannot be ruled out. Singh and Shimakawa [3.37] have derived separate effective masses of charge carriers in their extended and tail states. That means the density of states (DOS) of extended and tail states can be represented in two different parabolic forms. The relationship between the absorption coefficient and the distribution of electronic states for the case of a-Si:H may be found in [3.37, 42–44].

The existence of tail states in amorphous solids has a profound impact upon the band-to-band optical absorption. Unlike in a crystalline solid, the absorption of photons in an intrinsic amorphous solid can also occur for photon energies $\hbar\omega \leq E_0$ due to the presence of tail states in the forbidden gap. E_0 is the energy of the optical gap, which is usually close to the mobility gap – the energy difference between the conduction band and valence band mobility edges.

In a crystalline semiconductor, the energy and crystal momentum of an electron involved in an optical transition must be conserved. In an amorphous semiconductor, however, only the energy needs to be conserved [3.36, 37]. As a result, for optical transitions caused by photons of energy $\hbar\omega \geq E_0$ in amorphous semiconductors, the approach of a joint density of states is not applicable [3.37, 45]. One has to consider the product of the densities of both conduction and valence electronic states when calculating the absorption coefficient [3.37, 46]. Assuming that both the valence band and conduction band DOS functions have square-root dependencies on energy, one can derive the absorption coefficient α as [3.37]:

$$(\alpha\hbar\omega)^{1/2} = B^{1/2}(\hbar\omega - E_0), \quad (3.40)$$

where, if one assumes that the transition matrix element is constant,

$$B = \frac{1}{nc\epsilon_0} \left(\frac{e}{m_e^*} \right)^2 \left(\frac{L(m_e^*m_h^*)^{3/2}}{2^2\hbar^3} \right) \quad (3.41)$$

where m_e^* and m_h^* are the effective masses of an electron and a hole, respectively, L denotes the average interatomic separation in the sample, and n is the refractive index. Plotting $(\alpha\hbar\omega)^{1/2}$ as a function of $\hbar\omega$ yields a straight line that is usually referred to as Tauc's plot, from which one can determine the optical gap energy E_0 experimentally. Equation (3.40) is also known as Tauc's relation and B as Tauc's constant [3.47]. Experimental data from many, but not all, amorphous semiconductors and chalcogenides fit to (3.40) very well. Deviations from Tauc's relation have been observed. For instance, some experimental data in a-Si:H fit much better to a cubic root relation given by *Mott and Davis* [3.36]:

$$(\alpha\hbar\omega)^{1/3} = C(\hbar\omega - E_0) \quad (3.42)$$

and therefore the cubic root has been used to determine the optical gap E_0 . Here C is another constant.

If one considers that the optical transition matrix element is independent of photon energy [3.37, 48], one finds that the cubic root dependence on photon energy

can be obtained only when the valence band and conduction band DOS depend linearly on energy. Using such DOS functions, the cubic root dependence has been explained by *Mott and Davis* [3.36]. Another way to obtain the cubic root dependence has been suggested in [3.49]. Using (3.40), *Sokolov et al.* [3.49] have modeled the cubic root dependence on photon energy by considering the fluctuations in the optical band gap due to structural disorder. Although their approach gives a way of getting the cubic root dependence, *Sokolov et al.*'s model is hardly different from the linear density of states model suggested by *Mott and Davis* [3.36]. *Cody* [3.50] has shown an alternative approach using a photon energy-dependent transition matrix element. Thus, the absorption coefficient is obtained as [3.37]:

$$(\alpha\hbar\omega) = B'(\hbar\omega)^2(\hbar\omega - E_0)^2, \quad (3.43)$$

where

$$B' = \frac{e^2}{nc\epsilon_0} \left(\frac{(m_e^*m_h^*)^{3/2}}{2\pi^2\hbar^7\nu\rho_A} \right) Q_a^2, \quad (3.44)$$

and Q_a is the average separation between the excited electron and hole pair in an amorphous semiconductor, ν denotes the number of valence electrons per atom and ρ_A represents the atomic density per unit volume.

Equation (3.43) suggests that $(\alpha\hbar\omega)$ depends on the photon energy in the form of a polynomial of order 4. Then, depending on which term of the polynomial may be more significant in which material, one can get square, cubic or fourth root dependences of $(\alpha\hbar\omega)$ on



Fig. 3.10 Schematic illustration of the electronic energy states E_2 , E_c , E_{ct} , E_{vt} , E_v and E_{v2} in amorphous semiconductors. The shaded region represents the extended states. Energies E_2 and E_{v2} correspond to the centers of the conduction and valence extended states and E_{ct} and E_{vt} represent the ends of the conduction and valence tail states, respectively

the photon energy. In this case, (3.43) may be expressed as:

$$(\alpha\hbar\omega)^x \propto (\hbar\omega - E_0), \quad (3.45)$$

where $x \leq 1/2$. Thus, in a way, any deviation from the square root or Tauc's plot may be attributed to the energy-dependent matrix element [3.37, 46]. Another possible explanation has been recently discussed by Shimakawa and coworkers on the base of fractal theory [3.54].

Another problem is how to determine the constants B (3.41) and B' (3.44), which involve the effective masses of an electron and a hole. In other words, how do we determine the effective masses of charge carriers in amorphous solids? Recently, a simple approach [3.37, 46] has been developed to calculate the effective masses of charge carriers in amorphous solids. Different effective masses of charge carriers are obtained in the extended and tail states. The approach applies the concepts of tunneling and effective medium, and one obtains the effective mass of an electron in the conduction extended states, denoted by m_{ex}^* , and in the tail states, denoted by m_{et}^* , as:

$$m_{\text{ex}}^* \approx \frac{E_L}{2(E_2 - E_c)a^{1/3}} m_e, \quad (3.46)$$

and

$$m_{\text{et}}^* \approx \frac{E_L}{(E_c - E_{\text{ct}})b^{1/3}} m_e, \quad (3.47)$$

where:

$$E_L = \frac{\hbar^2}{m_e L^2}. \quad (3.48)$$

Here $a = \frac{N_1}{N} < 1$, N_1 is the number of atoms contributing to the extended states, $b = \frac{N_2}{N} < 1$, N_2 is the number of atoms contributing to the tail states, such that $a + b = 1$ ($N = N_1 + N_2$), and m_e is the free electron mass. The energy E_2 in (3.46) corresponds to the energy of the middle of the extended conduction states, at which the imaginary part of the dielectric constant becomes maximum (Fig. 3.10; see also Fig. 3.2).

Likewise, the effective masses of the hole m_{hx}^* and m_{ht}^* in the valence extended and tail states are obtained, respectively, as:

$$m_{\text{hx}}^* \approx \frac{E_L}{2(E_v - E_{v2})a^{1/3}} m_e, \quad (3.49)$$

and

$$m_{\text{ht}}^* \approx \frac{E_L}{(E_{vt} - E_v)b^{1/3}} m_e, \quad (3.50)$$

where E_{v2} and E_{vt} are energies corresponding to the half-width of the valence extended states and the end of the valence tail states, respectively; see Fig. 3.10.

Using (3.46) and (3.47) and the values of the parameters involved, different effective masses of an electron are obtained in the extended and tail states. Taking, for example, the density of weak bonds contributing to the tail states as 1 at. %, so $b = 0.01$ and $a = 0.99$, the effective masses and energies E_L calculated for hydrogenated amorphous silicon (a-Si:H) and germanium (a-Ge:H) are given in Table 3.6.

According to (3.46), (3.47), (3.49) and (3.50), for sp^3 hybrid amorphous semiconductors such as a-Si:H and a-Ge:H, effective masses of the electron and hole are expected to be the same. In these semiconductors, since the conduction and valence bands are two equal halves of the same electronic band, their widths are the same and that gives equal effective masses for the electron and the hole [3.37, 55]. This is one of the reasons for using $E_{\text{ct}} = E_{\text{vt}} = E_c/2$, which gives equal effective masses for electrons and holes in the tail states as well. This is different from crystalline solids where m_e^* and m_h^* are usually not the same. This difference between amorphous and crystalline solids is similar to, for example, having direct and indirect crystalline semiconductors but only direct amorphous semiconductors.

Using the effective masses from Table 3.6 and (3.41), B can be calculated for a-Si:H and a-Ge:H. The values thus obtained with the refractive index $n = 4$ for a-Si:H and a-Ge:H are $B = 6.0 \times 10^6 \text{ cm}^{-1} \text{ eV}^{-1}$ for a-Si:H and $B = 4.1 \times 10^6 \text{ cm}^{-1} \text{ eV}^{-1}$ for a-Ge:H, which are an order of magnitude higher than those estimated from

Table 3.6 Effective mass of electrons in the extended and tail states of a-Si:H and a-Ge:H calculated using Eqs. (3.46) and (3.47) for $a = 0.99$, $b = 0.01$ and $E_{\text{ct}} = E_{\text{vt}} = E_c/2$. E_L is calculated from (3.48). All energies are given in eV. Note that since the absorption coefficient is measured in cm^{-1} , the value used for the speed of light is in cm/s (a [3.51]; b [3.52]; c [3.33]; d [3.53])

	$L(\text{nm})$	E_2	E_c	E_L	$E_c - E_{\text{ct}}$	m_{ex}^*	m_{et}^*
a-Si:H	0.235 ^a	3.6 ^b	1.80 ^c	1.23	0.9	$0.34m_e$	$6.3m_e$
a-Ge:H	0.245 ^a	3.6	1.05 ^d	1.14	0.53	$0.22m_e$	$10.0m_e$

experiments [3.36]. However, considering the quantities involved in B (3.41), this can be regarded as a reasonable agreement.

In a recent paper, *Malik and O'Leary* [3.56] studied the distributions of conduction and valence band electronic states associated with a-Si:H. They noted that the effective masses associated with a-Si:H are material parameters that are yet to be experimentally determined. In order to remedy this deficiency, they fitted square-root DOS functions to experimental DOS data and found that $m_h^* = 2.34m_e$ and $m_e^* = 2.78m_e$.

The value of the constant B' in (3.44) can also be calculated theoretically, provided that Q_a is known. Using the atomic density of crystalline silicon and four valence electrons per atom, *Cody* [3.50] estimated $Q_a^2 = 0.9 \text{ \AA}^2$, which gives $Q_a \approx 0.095 \text{ nm}$, less than half the interatomic separation of 0.235 nm in a-Si:H, but of the same order of magnitude. Using $\nu = 4$, $\rho_A = 5 \times 10^{28} \text{ m}^{-3}$, $Q_a^2 = 0.9 \text{ \AA}^2$, and extended state effective masses, we get $B' = 4.6 \times 10^3 \text{ cm}^{-1} \text{ eV}^{-3}$ for a-Si:H and $1.3 \times 10^3 \text{ cm}^{-1} \text{ eV}^{-3}$ for a-Ge:H. *Cody* estimated an optical gap, $E_0 = 1.64 \text{ eV}$, for a-Si:H, which, using in (3.43), gives $\alpha = 1.2 \times 10^3 \text{ cm}^{-1}$ at a photon energy of $\hbar\omega = 2 \text{ eV}$. This agrees reasonably well with the $\alpha = 6.0 \times 10^2 \text{ cm}^{-1}$ used by *Cody*. If we use interatomic spacing in place of Q_a in (3.44), we get $B' = 2.8 \times 10^4 \text{ cm}^{-1} \text{ eV}^{-3}$, and then the corresponding

absorption coefficient becomes $3.3 \times 10^3 \text{ cm}^{-1}$. This suggests that, in order to get an estimate, one may use the interatomic spacing in place of Q_a if the latter is unknown. Thus, both B and B' can be determined theoretically, a task not possible before due to a lack of knowledge of the effective masses in amorphous semiconductors.

The absorption of photons of energy less than the band gap energy, $\hbar\omega < E_0$ in amorphous solids involves the localized tail states and hence follows neither (3.40) nor (3.42). Instead, the absorption coefficient depends on the photon energy exponentially, as given in (3.39), giving rise to Urbach's tail. *Abe and Toyozawa* [3.57] have calculated the interband absorption spectra in crystalline solids, introducing the Gaussian site diagonal disorder and applying the coherent potential approximation. They have shown that Urbach's tail occurs due to static disorder (structural disorder). However, the current stage of understanding is that Urbach's tail in amorphous solids occurs due to both thermal effects and static disorder [3.50]. More recent issues in this area have been addressed by *Orapunt and O'Leary* [3.48].

Keeping the above discussion in mind, the optical absorption in amorphous semiconductors near the absorption edge is usually characterized by three types of optical transitions corresponding to transitions between tail and tail states, tail and extended states, and extended and extended states. The first two types correspond to $\hbar\omega \leq E_0$, and the third one corresponds to $\hbar\omega \geq E_0$. Thus, the plot of absorption coefficient versus photon energy (α versus $\hbar\omega$) has three different regions, A, B and C, respectively, that correspond to these three characteristic optical transitions shown in Fig. 3.11.

In the small absorption coefficient range A (also called the weak absorption tail, WAT), where $\alpha < 10^{-1} \text{ cm}^{-1}$, the absorption is controlled by optical transitions from tail-to-tail states. As stated above, the localized tail states in amorphous semiconductors are contributed by defects. To some extent, the absolute value of absorption in region A may be used to estimate the density of defects in the material. In region B, where typically $10^{-1} < \alpha < 10^4 \text{ cm}^{-1}$, the absorption is related to transitions from the localized tail states above the valence band edge to extended states in the conduction band and/or from extended states in the valence band to localized tail states below the conduction band. The spectral dependence of α usually follows the so-called Urbach rule, given in (3.39). For many amorphous semiconductors, ΔE has been related to the width of the valence (or conduction) band tail states, and may be used to compare the “widths” of such localized tail states in

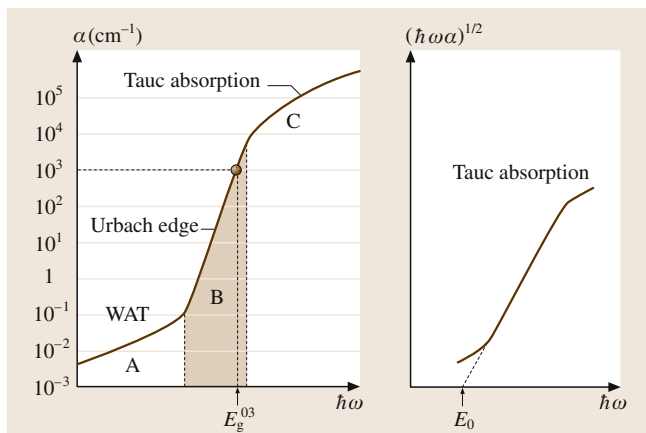


Fig. 3.11 Typical spectral dependence of the optical absorption coefficient in amorphous semiconductors. In the A and B regions, the absorption is controlled by optical transitions between tail and tail states and tail and extended states, respectively, and in the C region it is dominated by transitions from extended to extended states. In domain B, the absorption coefficient follows the Urbach rule (3.39). In region C, the absorption coefficient follows Tauc's relation (3.40) in a-Si:H, as shown in the figure on the right-hand side

Table 3.7 Optical bandgaps of a-Si_{1-x}C_x:H films obtained from Tauc's (3.40), Sokolov et al.'s (3.42) and Cody's (3.43) relations [3.58]

	E_g at $\alpha = 10^3 \text{ cm}^{-1}$	E_g at $\alpha = 10^4 \text{ cm}^{-1}$	$E_g = E_0$ (Tauc)	$E_g = E_0$ (Cody)	$E_g = E_0$ (Sokolov)	ΔE (meV)
a-Si:H	1.76	1.96	1.73	1.68	1.60	46
a-Si _{0.88} C _{0.18} :H	2.02	2.27	2.07	2.03	1.86	89

different materials; ΔE is typically 0.05–0.1 eV. In region C, the absorption is controlled by transitions from extended to extended states. For many amorphous semiconductors, the α versus $\hbar\omega$ behavior follows the Tauc relation given in (3.40). The optical bandgap, E_0 , determined for a given material from the α versus $\hbar\omega$ relations obtained in (3.40), (3.42) and (3.43), can vary as shown in Table 3.7 for a-Si:H alloys.

3.3.4 Exciton Absorption

Excitons in Crystalline Semiconductors

Optical absorption in crystalline semiconductors and insulators can create an exciton, which is an electron and hole pair excited by a photon and bound together through their attractive Coulomb interaction. This means that the absorbed optical energy remains held within the solid for the lifetime of the exciton. There are two types of excitons that can be formed in nonmetallic solids: Wannier or Wannier–Mott excitons and Frenkel excitons. The concept of Wannier–Mott excitons is valid for inorganic semiconductors like Si, Ge and GaAs, because in these materials the large overlap of interatomic electronic wavefunctions enables the electrons and holes to be far apart but bound in an excitonic state. For this reason, these excitons are also called large-radii orbital excitons. Excitons formed in organic crystals are called Frenkel excitons. In organic semiconductors/insulators or molecular crystals, the intermolecular separation is large and hence the overlap of intermolecular electronic wavefunctions is very small and electrons remain tightly bound to individual molecules. Therefore, the electronic energy bands are very narrow and closely related to individual molecular electronic energy levels. In these solids, the absorption of photons occurs close to the individual molecular electronic states and excitons are also formed within the molecular energy levels (see [3.59]). Such excitons are therefore also called molecular excitons. For a more detailed look at the theory of Wannier and Frenkel excitons, readers may like to refer to Singh [3.59].

In Wannier–Mott excitons, the Coulomb interaction between the hole and electron can be viewed as an

effective hydrogen atom with, for example, the hole establishing the coordinate reference frame about which the reduced mass electron moves. If the effective masses of the isolated electron and hole are m_e^* and m_h^* , respectively, their reduced mass μ_x is given by:

$$\mu_x^{-1} = (m_e^*)^{-1} + (m_h^*)^{-1}. \quad (3.51)$$

Note that in the case of so-called hydrogenic impurities in semiconductors (both shallow donor and acceptor impurities), the reduced mass of the nucleus takes the place of one of the terms in (3.51) and hence the reduced mass is given to good approximation by the effective mass of the appropriate carrier. When the exciton is the carrier, the effective masses are comparable and hence the reduced mass is markedly lower; accordingly, the exciton binding energy is markedly lower than that for hydrogenic impurities. The energy of a Wannier–Mott exciton is given by (e.g. [3.59]):

$$E_x = E_g + \frac{\hbar^2 K^2}{2M} - E_n, \quad (3.52)$$

where E_g is the band gap energy, $\hbar K$ is the linear momentum, $M (= m_e^* + m_h^*)$ is the effective mass associated with the center of mass of an exciton, and E_n is the exciton binding energy given by

$$E_n = \frac{\mu_x e^4 \kappa^2}{2\hbar^2 \epsilon^2} \frac{1}{n^2} = \frac{R_y^x}{n^2}, \quad (3.53)$$

where e is the electronic charge, $\kappa = \frac{1}{4\pi\epsilon_0}$, ϵ is the static dielectric constant of the solid, and n is the principal quantum number associated with the internal excitonic states $n = 1(s), 2(p), \dots$ (see Fig. 3.12). According to (3.53), the excitonic states are formed within the band gap below the conduction band edge. However, as the exciton binding energy is very small (a few meV in bulk Si and Ge crystals), exciton absorption peaks can only be observed at very low temperatures. R_y^x is the so-called effective Rydberg of the exciton given by $R_y(\mu_x/m_e)/\epsilon^2$, where $R_y = 13.6 \text{ eV}$. For bulk GaAs, the exciton binding energy ($n = 1$) corresponds to about 5 meV. Extrapolating from the hydrogen atom model, the extension of the excitonic wavefunction can be found from an effective Bohr radius a_B^* given in terms of the Bohr radius

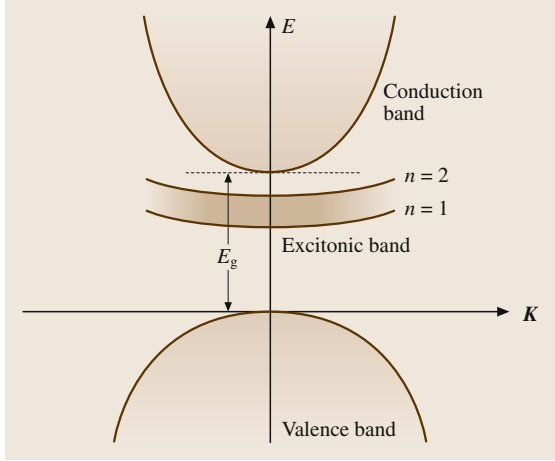


Fig. 3.12 Schematic illustration of excitonic bands for $n = 1$ and 2 in semiconductors. E_g represents the energy gap

as $a_B = \hbar^2/\pi e^2$; that is, $a_B^* = a_B(\epsilon/\mu_x)$. For GaAs this corresponds to about 12 nm or about 21 lattice constants – in other words the spherical volume of the exciton radius contains $\approx (a_B^*/a_0)^3$ or ≈ 9000 unit cells, where a_0 is the lattice constant of GaAs. As $Ry^* \ll E_g$ and $a_B^* \gg a_0$, it is clear from this example that the excitons in GaAs are large-radii orbital excitons, as stated above. It should be noted that the binding energies of excitons in semiconductors tend to be a strong function of the bandgap. The dependencies of R_y^x (exciton binding energy) and a_B^*/a_0 on the semiconductor bandgap are shown in Fig. 3.13a and b, respectively. For excitons with large binding energies and correspondingly small radii (approaching the size of a lattice parameter), the excitons become localized on a lattice site, as observed in most organic semiconductors. As stated above, such excitons are commonly referred to as Frenkel excitons or molecular excitons. Unlike Wannier-based excitons which are typically dissociated at room temperature, these excitations are stable at room temperature. For the binding energies of Frenkel excitons, one can refer to Singh for example [3.59].

Excitons can recombine radiatively, emitting a series of hydrogen-like spectral lines, as described by (3.52). In bulk (3-D) semiconductors such as GaAs, exciton lines can only be observed at low temperatures: they are easily dissociated by thermal fluctuations. On the other hand, in quantum wells and other structures of reduced dimensionality, the spatial confinement of both the electron and hole wavefunctions in the same layer ensures strong excitonic transitions of a few meV be-

low the bandgap, even at room temperature. Excitonic absorption is well-located spectrally and very sensitive to optical saturation. For this reason, it plays an important role in nonlinear semiconductor devices (nonlinear Fabry–Perot resonators, nonlinear mirrors, saturable absorbers, and so on). If the valence hole is a heavy hole, the exciton is called a heavy exciton; conversely, if the valence hole is light, the exciton is a light exciton. For practical purposes, the excitonic contribution to the overall susceptibility around the resonance frequency ν_{ex} can be written as:

$$\chi_{exc} = -A_0 \frac{(\nu - \nu_{ex}) + i\Gamma_{ex}}{(\nu - \nu_{ex})^2 + \Gamma_{ex}^2(1 + S)}, \quad (3.54)$$

where Γ_{ex} is the linewidth and $S = I/I_S$ the saturation parameter of the transition. For instance, in GaAs MQW, the saturation intensity I_S is as low as 1 kW/cm², and Γ_{ex} (≈ 3.55 meV at room temperature) varies with the temperature according to

$$\Gamma_{ex} = \Gamma_0 + \frac{\Gamma_1}{\exp(\hbar\omega_{LO}/kT) - 1}, \quad (3.55)$$

where $\hbar\Gamma_0$ is the inhomogeneous broadening (≈ 2 meV), $\hbar\Gamma_1$ is the homogeneous broadening (≈ 5 meV), and

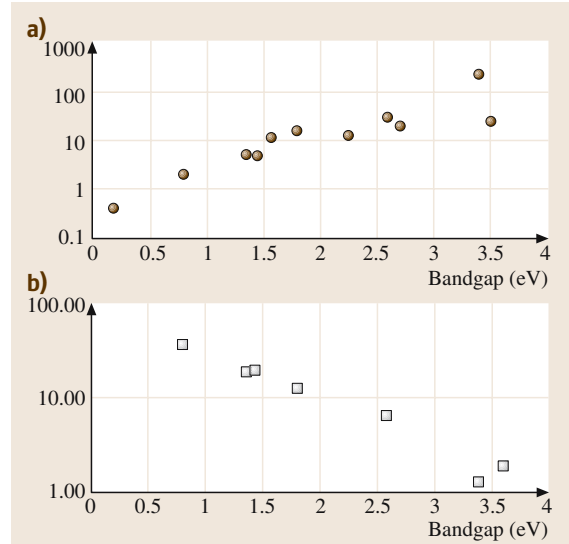


Fig. 3.13a,b Dependence of the (a) exciton binding energy (R_y^x) (3.29) and (b) size (in terms of the ratio of the excitonic Bohr radius to lattice constant (a_B^*/a_0)) as a function of the semiconductor bandgap. Exciton binding energy increases along with a marked drop in exciton spreading as bandgap increases. The Wannier-based description is not appropriate above a bandgap of about 2 eV

$\hbar\omega_{\text{LO}}$ is the longitudinal optical phonon energy (≈ 36 meV). At high carrier concentrations (provided either by electrical pumping or by optical injection), the screening of the Coulomb attractive potential by free electrons and holes provides an efficient mechanism for saturating the excitonic line.

The above discussion refers to so-called free excitons formed between conduction band electrons and valence band holes. According to (3.52), such an excitation is able to move throughout a material with a given center-of-mass kinetic energy (second term on the right hand side). It should be noted, however, that free electrons and holes move with a velocity $\hbar(dE/dk)$ where the derivative is taken for the appropriate band edge. To move through a crystal, both the electron and the hole must have identical translational velocities, restricting the regions in k -space where these excitations can occur to those with $(dE/dk)_{\text{electron}} = (dE/dk)_{\text{hole}}$, commonly referred to as critical points.

A number of more complex pairings of carriers can also occur, which may also include fixed charges or ions. For example, for the case of three charged entities with one being an ionized donor impurity (D+), the following possibilities can occur: (D+)(+)(-), (D+)(-)(-) and (+)(+)(-) as excitonic ions, and (+)(+)(-)(-) and (D+)(+)(-)(-) as biexcitons or even bigger excitonic molecules (see [3.60]). Complexity abounds in these systems, as each electronic level possesses a fine structure corresponding to allowed rotational and vibrational levels. Moreover, the effective mass is often anisotropic. Note that when the exciton or exciton complex is bound to a fixed charge, such as an ionized donor or acceptor center in the material, the exciton or exciton complex is referred to as a bound exciton. Indeed, bound excitons may also involve neutral fixed impurities. It is usual to relate the exciton in these cases to the center binding them; thus, if an exciton is bound to a donor impurity, it is usually termed a donor-bound exciton.

Excitons in Amorphous Semiconductors

The concept of excitons is traditionally valid only for crystalline solids. However, several observations in the photoluminescence spectra of amorphous semiconductors have revealed the occurrence of photoluminescence associated with singlet and triplet excitons [3.37]. Applying the effective mass approach, a theory for the Wannier–Mott excitons in amorphous semiconductors has recently been developed in real coordinate space [3.37, 46, 55, 61]. The energy of an exciton thus

derived is obtained as:

$$W_x = E_0 + \frac{P^2}{2M} - E_n(S), \quad (3.56)$$

where P is the linear momentum associated with the exciton's center of motion and $E_n(S)$ is the binding energy of the exciton, given by

$$E_n(S) = \frac{\mu_x e^4 \kappa^2}{2\hbar^2 \varepsilon'(S)^2 n^2}, \quad (3.57)$$

where

$$\varepsilon'(S) = \varepsilon \left[1 - \frac{(1-S)}{A} \right]^{-1}, \quad (3.58)$$

where S is the spin (S being = 0 for singlet and = 1 for triplet) of an exciton and A is a material-dependent constant representing the ratio of the magnitude of the Coulomb and exchange interactions between the electron and the hole of an exciton. Equation (3.57) is analogous to (3.53) obtained for excitons in crystalline solids for $S = 1$. This is because (3.53) is derived within the large-radii orbital approximation, which neglects the exchange interaction and hence is valid only for triplet excitons [3.59, 62]. As amorphous solids lack long-range order, the exciton binding energy is found to be larger in amorphous solids than in their crystalline counterparts; for example, the binding energy is higher in hydrogenated amorphous silicon (a-Si:H) than in crystalline silicon (c-Si). This is the reason that it is possible to observe both singlet and triplet excitons in a-Si:H [3.63] but not in c-Si.

Excitonic Absorption

Since exciton states lie below the edge of the conduction band in a crystalline solid, absorption to excitonic states is observed below this edge. According to (3.53), the difference in energy in the bandgap and the excitonic absorption gives the binding energy. As the exciton–photon interaction operator and excited electron and hole pair and photon interaction operator depend only on their relative motion, these interactions take the same form for band-to-band and excitonic absorption. Therefore, to calculate the excitonic absorption coefficient, one can use the same form of interaction as that used for band-to-band absorption, but one must use the joint density of states. Using the joint density of states, the absorption coefficient associated with the excitonic states in crystalline semiconductors is obtained as ([3.37]):

$$\alpha \hbar\omega = A_x (\hbar\omega - E_x)^{1/2} \quad (3.59)$$

with the constant $A_x = 4\sqrt{2} e^2 |p_{xv}|^2 / nc \sqrt{\mu_x} \hbar^2$, where p_{xv} is the transition matrix element between the valence and excitonic bands. Equation (3.59) is similar to that seen for direct band-to-band transitions, discussed above ((3.60)), and is only valid for the photon energies $\hbar\omega \geq E_x$. There is no absorption below the excitonic ground state in pure crystalline solids. Absorption of photons to excitonic energy levels is possible through either the excitation of electrons to higher energy levels in the conduction band and then nonradiative relaxation to the excitonic energy level, or through the excitation of an electron directly to the exciton energy level. Excitonic absorption occurs in both direct and indirect semiconductors.

In amorphous semiconductors, the excitonic absorption and photoluminescence can be quite complicated. According to (3.56), the excitonic energy level is below the optical band gap by an energy equal to the binding energy given in (3.57). However, there are four transition possibilities: (i) extended valence to extended conduction states, (ii) valence to extended conduction states, (iii) valence extended to conduction tail states, and (iv) valence tail to conduction tail states. These possibilities will have different optical gap energies, E_0 , and different binding energies. Transition (i) will give rise to absorption in the free exciton states, transitions (ii) and (iii) will give absorption in the bound exciton states, because one of the charge carriers is localized in the tail states, and absorption through transition (iv) will create localized excitons, which are also called geminate pairs. This can be visualized as follows: if an electron–hole pair is excited by a high-energy photon through transition (i) and forms an exciton, initially its excitonic energy level and the corresponding Bohr radius will have a reduced mass corresponding to both charge carriers being in extended states. As such an exciton relaxes downward nonradiatively, its binding energy and excitonic Bohr radius will change because its effective mass changes in the tail states. When both charge carriers reach the tail

Fig. 3.14 Energy level diagram of the low-lying $4f^N$ states of trivalent ions doped in LaCl_3 . After [3.64–66]. The pendant semicircles indicate fluorescent levels ►

states (transition (iv)), their excitonic Bohr radius will be maintained although they are localized.

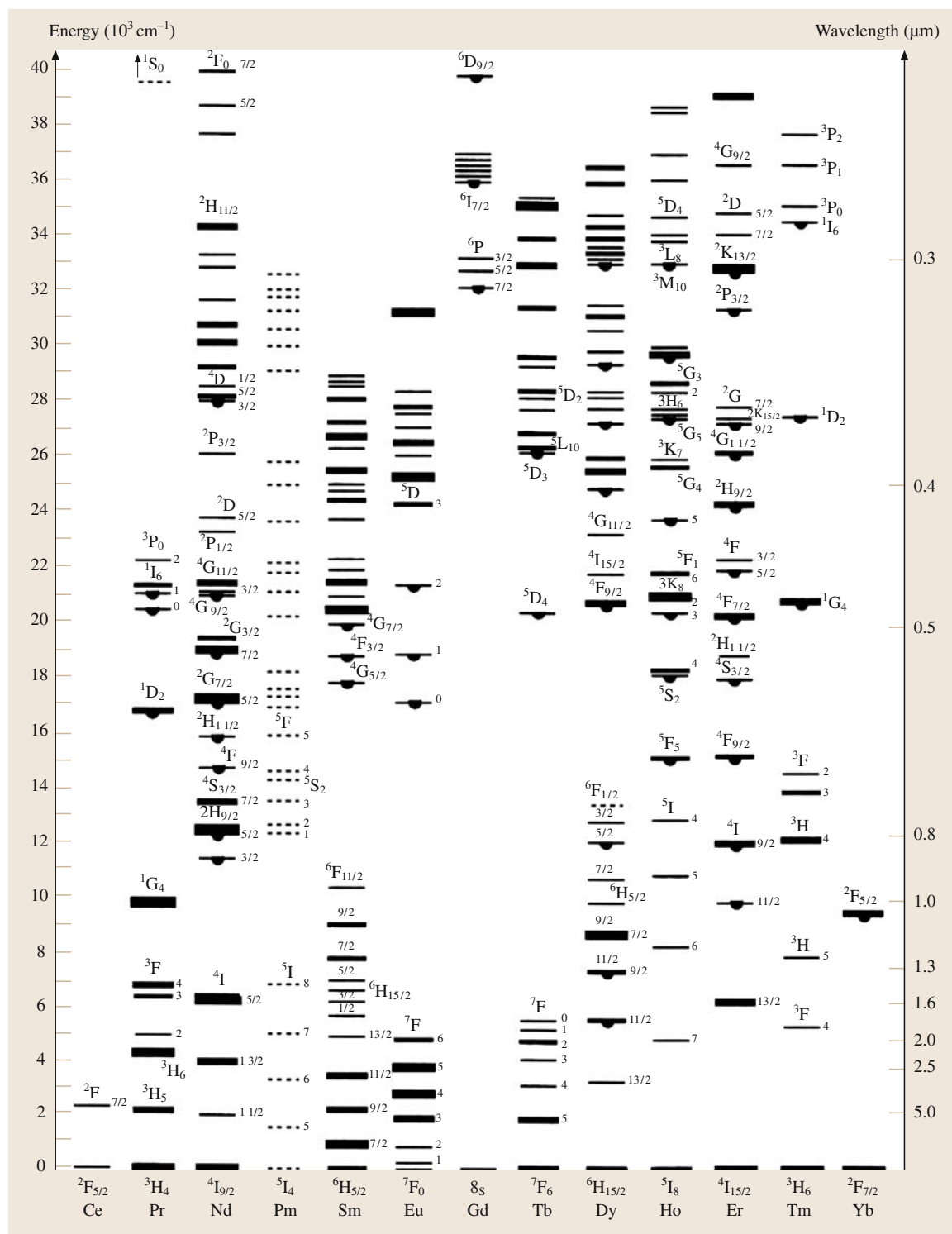
The excitonic absorption coefficient in amorphous semiconductors can be calculated using the same approach as presented in Sect. 3.3.3, and similar expressions to (3.40) and (3.43) are obtained. This is because the concept of the joint density of states is not applicable in amorphous solids. Therefore, by replacing the effective masses of the charge carriers by the excitonic reduced mass and the distance between the excited electron and hole by the excitonic Bohr radius, one can use (3.40) and (3.43) to calculate the excitonic absorption coefficients for the four possible transitions above in amorphous semiconductors. However, such a detailed calculation of the excitonic transitions in amorphous semiconductors is yet to be performed.

3.3.5 Impurity Absorption

Impurity absorption can be observed as the absorption coefficient peaks lying below fundamental (band-to-band) and excitonic absorption. It is usually related to the presence of ionized impurities or, simply, ions. The peaks occur due to electronic transitions between ionic electronic states and the conduction/valence band or due to intra-ionic transitions (within d or f shells, between s and d shells, and so on). The first case leads to intense and broad lines, while the characteristics of the features arising from the latter case depend on whether or not these transitions are allowed by parity selection rules. For allowed transitions, the absorption peaks are quite intense and broad, while forbidden transitions produce weak and narrow peaks. General reviews of this topic may be found in *Blasse* and *Grabmaier* [3.67], *Hen-*

Table 3.8 Occupation of outer electronic shells for rare earth elements

57	La	4s ²	4p ²	4d ¹⁰	–	5s ²	5p ⁶	5d ¹	6s ²
58	Ce	4s ²	4p ²	4d ¹⁰	4f ¹	5s ²	5p ⁶	5d ¹	6s ²
59	Pr	4s ²	4p ²	4d ¹⁰	4f ³	5s ²	5p ⁶	–	6s ²
60	Nd	4s ²	4p ²	4d ¹⁰	4f ⁴	5s ²	5p ⁶	–	6s ²
...									
68	Er	4s ²	4p ²	4d ¹⁰	4f ¹²	5s ²	5p ⁶	–	6s ²
...									
70	Yb	4s ²	4p ²	4d ¹⁰	4f ¹⁴	5s ²	5p ⁶	–	6s ²
71	Lu	4s ²	4p ²	4d ¹⁰	4f ¹⁴	5s ²	5p ⁶	5d ¹	6s ²



dereson and Imbusch [3.65] and DiBartolo [3.68]. In the following section, we concentrate primarily on the properties of rare earth ions, which are of great importance in modern optoelectronics.

Optical Absorption of Trivalent Rare Earth Ions: Judd–Ofelt Analysis

Rare earths (REs) is the common name used for the elements from Lanthanum (La) to Lutetium (Lu). They have atomic numbers of 57 to 71 and form a separate group in Periodic Table. The most notable feature of these elements is an incompletely filled 4f shell. The electronic configurations of REs are listed in Table 3.8. The RE may be embedded in different host materials in the form of divalent or trivalent ions. As divalent ions, REs exhibit broad absorption–emission lines related to allowed 4f→5d transitions. In trivalent form, REs lose two 6s electrons and one 4f or 5d electron. The Coulomb interaction of a 4f electron with a positively charged core means that the 4f level gets split into complicated set of manifolds with energies, to a first approximation, that are virtually independent of the host matrix because the 4f level is well screened from external influences by the 5s and 5p shells [3.69]. The Fig. 3.14 shows an energy level diagram for the low-lying 4f^N states of the trivalent ions embedded in LaCl₃. To a second approximation, the exact construction and precise energies of the manifolds depend on the host material, via crystal field and via covalent interactions with the ligands surrounding the RE ion. A ligand is an atom (or molecule or radical or ion) with one or more unshared pairs of electrons that can attach to a central metallic ion (or atom) to form a coordination complex. Examples of ligands include ions (F[−], Cl[−], Br[−], I[−], S^{2−}, CN[−], NCS[−], OH[−], NH₂[−]) and molecules (NH₃, H₂O, NO, CO) that donate a pair of electrons to a metal atom or ion. Some ligands that share electrons with metals form very stable complexes.

Optical transitions between 4f manifold levels are forbidden by a parity selection rule which states that the wavefunctions of the initial and final states of an atomic (ionic) transition must have different parities for it to be permitted. Parity is a property of any function (or quantum mechanical state) that describes the function after mirror reflection. Even functions (states) are symmetric (identical after reflection, for example a cosine function), while odd functions (states) are antisymmetric (for example a sine function). The parity selection rule may be partially removed for an ion (or atom) embedded in host material due to the action of the crystal field, which gives rise to “forbidden lines”. The crystal

field is the electric field created by a host material at the position of the ion.

The parity selection rule is slightly removed by the admixture of 5d states with 4f states and by the disturbed RE ion symmetry due to the influence of the host, which increases with the covalency. Higher covalency implies stronger sharing of electrons between the RE ion and the ligands. This effect is known as the nephelauxetic effect. The resulting absorption–emission lines are characteristic of individual RE ions and quite narrow because they are related to forbidden inner shell 4f transitions.

Judd–Ofelt (JO) analysis allows the oscillator strength of an electric dipole (ED) transition between two states of a trivalent rare earth (RE) ion embedded in a particular host lattice to be calculated. The possible states of an RE ion are often referred to as ^{2S+1}L_J, where L = 0, 1, 2, 3, 4, 5, 6 . . . determines the electron’s total angular momentum, and is conventionally represented by the letters S, P, D, F, G, I. The term (2S + 1) is called the spin multiplicity and represents the number of spin configurations, while J is the total angular momentum, which is the vector sum of the overall (total) angular momentum and the overall spin (**J** = **L** + **S**). The value (2J + 1) is called the multiplicity and corresponds to the number of possible combinations of overall angular momentum and overall spin that yield the same J. Thus, the notation ⁴I_{15/2} for the ground state of Er³⁺ corresponds to the term (J, L, S) = (15/2, 6, 3/2), which has a multiplicity of 2J + 1 = 16 and a spin multiplicity of 2S + 1 = 4. If the wavefunctions |ψ_i⟩ and |ψ_f⟩ correspond to the initial (^{2S+1}L_J) and final (^{2S'+1}L'_{J'}) states of an electric dipole transition of an RE ion, the line strength of this transition, according to JO theory, can be calculated using:

$$S_{\text{ed}} = |\langle \psi_f | H_{\text{ed}} | \psi_i \rangle|^2 = \sum_{k=2,4,6} \Omega_k \left| \left\langle f_{\gamma}^N S' L' J' \left| U^{(k)} \right| f_{\gamma}^N S L J \right\rangle \right|^2, \quad (3.60)$$

where H_{ed} is the ED interaction Hamiltonian, Ω_k are coefficients reflecting the influence of the host material, and $U^{(k)}$ are reduced tensor operator components, which are virtually independent of the host material, and their values are calculated using the so-called intermediate coupling approximation (see [3.70]). The theoretical values of S_{ed} calculated from this are compared with the values derived from experimental data using

$$S_{\text{exp}} = \frac{3hcn}{8\pi^3 e^2 \langle \lambda \rangle} \frac{2J+1}{\chi_{\text{ed}}} \int_{\text{band}} \frac{\alpha(\lambda)}{\rho} d\lambda, \quad (3.61)$$

where $\langle\lambda\rangle$ is the mean wavelength of the transition, h is the Planck constant, c is the speed of light, e is the elementary electronic charge, $\alpha(\lambda)$ is the absorption coefficient, ρ is the RE ion concentration, n is the refraction index and the factor $\chi_{\text{ed}} = (n^2 + 2)^2/9$ is the so-called local field correction. The key idea of JO analysis is to minimize the discrepancy between experimental and calculated values of line strength, S_{ed} and S_{exp} , by choosing the coefficients Ω_k , which are used to characterize and compare materials, appropriately. The complete analysis should also include the magnetic dipole transitions [3.71]. The value of Ω_2 is of prime importance because it is the most sensitive to the local structure and material composition and is correlated with the degree of covalence. The values of Ω_k are used to calculate radiative transition probabilities and appropriate radiative lifetimes of excited states, which are very useful for numerous optical applications. More detailed analysis may be found in, for example, [3.71]. Ω_k values for different ions and host materials can be found in *Gschneidner, Jr. and Eyring* [3.72].

3.3.6 Effects of External Fields

Electroabsorption and the Franz–Keldysh Effect

Electroabsorption is the absorption of light in a device where the absorption is induced by an applied (or changing) electric field within the device. Such a device is an electroabsorption modulator. There are three fundamental types of electroabsorption processes. In the Franz–Keldysh process, a strong applied field modifies the photon-assisted probability of an electron tunneling from the valence band to the conduction band, and thus it corresponds to an effective reduction in the “bandgap energy”, inducing the absorption of light with photon energies of slightly less than the bandgap. It was first observed for CdS, in which the absorption edge was observed to shift to lower energies with the applied field; that is, photon absorption shifts to longer wavelengths with the applied field. The effect is normally quite small but is nonetheless observable. In this type of electroabsorption modulation, the wavelength is typically chosen to be slightly smaller than the bandgap wavelength so that absorption is negligible. When a field is applied, the absorption is enhanced by the Franz–Keldysh effect. In free carrier absorption, the concentration of free carriers N in a given band is changed (modulated), for example, by an applied voltage, changing the extent of photon absorption. The absorption coefficient is proportional to N and to

the wavelength λ of the light raised to some power, typically 2–3. In the confined Stark effect, the applied electric field modifies the energy levels in a quantum well. The energy levels are reduced by the field by an amount proportional to the square of the applied field. A multiple quantum well (MQW) pin-type device has MQWs in its intrinsic layer. Without any applied bias, light with photon energy just less than the QW exciton excitation energy will not be significantly absorbed. When a field is applied, the energy levels are lowered and the incident photon energy is then sufficient to excite an electron and hole pair in the QWs. The relative transmission decreases with the reverse bias V_r applied to the pin device. Such MQW pin devices are usually not very useful in the transmission mode because the substrate material often absorbs the light (for example a GaAs/AlGaAs MQW pin would be grown on a GaAs substrate, which would absorb the radiation that excites the QWs). Thus, a reflector would be needed to reflect the light back before it reaches the substrate; such devices have indeed been fabricated.

The Faraday Effect

The Faraday effect, originally observed by Michael Faraday in 1845, is the rotation of the plane of polarization of a light wave as it propagates through a medium subjected to a magnetic field parallel to the direction of propagation of the light. When an optically inactive material such as glass is placed in a strong magnetic field and plane-polarized light is sent along the direction of the magnetic field, the emerging light’s plane of polarization is rotated. The magnetic field can be applied, for example, by inserting the material into the core of a magnetic coil – a solenoid. The specific rotatory power induced, given by θ/L , has been found to be proportional to the magnitude of the applied magnetic field B , which gives the amount of rotation as:

$$\theta = \vartheta BL, \quad (3.62)$$

where L is the length of the medium, and ϑ is the so-called Verdet constant, which depends on the material and the wavelength of the light. The Faraday effect is typically small. For example, a magnetic field of ≈ 0.1 T causes a rotation of about 1° through a glass rod of length 20 mm. It appears that “optical activity” is induced by the application of a strong magnetic field to an otherwise optically inactive material. There is, however, an important distinction between natural optical activity and the Faraday effect. The sense of rotation θ observed in the

Table 3.9 Verdet constants for some materials

Material	Quartz 589.3nm	Flint glass 632nm	Tb–Ga garnet 632nm	Tb–Ga garnet 1064nm	ZnS 589.3nm	Crown glass 589.3	NaCl 589.3
ϑ (rad m ⁻¹ T ⁻¹)	4.0	4.0	134	40	82	6.4	9.6

Faraday effect for a given material (Verdet constant) depends only on the direction of the magnetic field \mathbf{B} . If ϑ is positive, for light propagating parallel to \mathbf{B} , the optical field \mathbf{E} rotates in the same sense as an advancing right-handed screw pointing in the direction of \mathbf{B} . The direction of light propagation does not change the absolute sense of rotation of θ . If we reflect the wave to pass through the medium again, the rotation increases to

2θ . The Verdet constant depends not only on the wavelength λ but also on the charge-to-mass ratio of the electron and the refractive index $n(\lambda)$ of the medium through:

$$\vartheta = -\frac{(e/m_e)}{2c} \lambda \frac{dn}{d\lambda}. \quad (3.63)$$

Verdet constants for some glasses are listed in Table 3.9.

3.4 Thin Film Optics

Thin film optics involves multiple reflections of light entering a thin film dielectric (typically on a substrate) so that the reflection and transmission coefficients are determined by multiple wave interference phenomena, as shown in Fig. 3.15.

Consider a thin film coated on a substrate. Assuming that the incident wave has an amplitude of A_0 , then there are various transmitted and reflected waves, as shown in Fig. 3.15. We then have the following amplitudes based on the definitions of the reflection and transmission coefficients:

$$\begin{aligned}
 A_1 &= A_0 r_{12} \\
 A_2 &= A_0 t_{12} r_{23} t_{21} \\
 A_3 &= A_0 t_{12} r_{23} r_{21} r_{23} t_{21} \\
 B_1 &= A_0 t_{12} \\
 B_2 &= A_0 t_{12} r_{23} \\
 B_3 &= A_0 t_{12} r_{23} r_{21} \\
 B_4 &= A_0 t_{12} r_{23} r_{21} r_{23} \\
 B_5 &= A_0 t_{12} r_{23} r_{21} r_{23} r_{21} \\
 B_6 &= A_0 t_{12} r_{23} r_{21} r_{23} r_{21} r_{23} \\
 C_1 &= A_0 t_{12} t_{23} \\
 C_2 &= A_0 t_{12} r_{23} r_{21} t_{23} \\
 C_3 &= A_0 t_{12} r_{23} r_{21} r_{23} r_{21} t_{23}
 \end{aligned} \quad (3.64)$$

and so on, where r_{12} is the reflection coefficient of a wave in medium 1 incident on medium 2, and t_{12} is the transmission coefficient from medium 1 into 2. For simplicity, we will assume normal incidence. The phase change upon traversing the thin film thickness d is $\phi = (2\pi/\lambda)n_2d$, where λ is the free space wavelength.

The wave must be multiplied by $\exp(i\phi)$ to account for this phase difference.

The reflection and transmission coefficients are given by

$$\begin{aligned}
 r_1 &= r_{12} = \frac{n_1 - n_2}{n_1 + n_2} = -r_{21}, \\
 r_2 &= r_{23} = \frac{n_2 - n_3}{n_2 + n_3},
 \end{aligned} \quad (3.65a)$$

and

$$\begin{aligned}
 t_1 &= t_{12} = \frac{2n_1}{n_1 + n_2}, \quad t_2 = t_{21} = \frac{2n_2}{n_1 + n_2}, \\
 t_3 &= t_{23} = \frac{2n_3}{n_2 + n_3},
 \end{aligned} \quad (3.65b)$$

where

$$1 - t_1 t_2 = r_1^2. \quad (3.66)$$

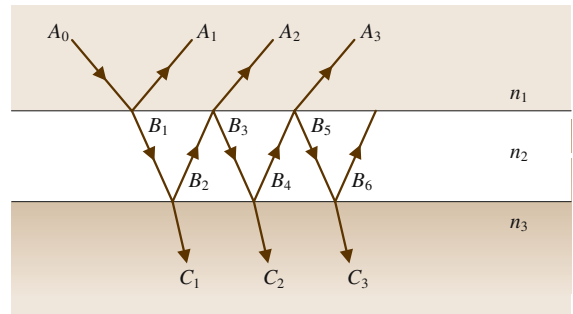


Fig. 3.15 Thin film coated on a substrate and multiple reflections of incident light, where n_1, n_2 and n_3 are the refractive indices of the medium above the thin film, the thin film, and the substrate, respectively

The total reflection coefficient is then obtained as:

$$r = \frac{A_{\text{reflected}}}{A_0} = r_1 - \frac{t_1 t_2}{r_1} \sum_{k=1}^{\infty} \left(-r_1 r_2 e^{-i2\phi} \right)^k, \quad (3.67)$$

which can be summed to be:

$$r = \frac{r_1 + r_2 e^{-i2\phi}}{1 + r_1 r_2 e^{-i2\phi}}. \quad (3.68)$$

The total transmission coefficient is obtained as:

$$\begin{aligned} t = \frac{C_{\text{transmitted}}}{A_0} &= -\frac{t_1 t_{23} e^{i\phi}}{r_1 r_2} \sum_{k=1}^{\infty} \left(-r_1 r_2 e^{-i2\phi} \right)^k \\ &= \left(\frac{t_1 t_3 e^{i\phi}}{r_1 r_2} \right) \frac{r_1 r_2 e^{-i2\phi}}{1 + r_1 r_2 e^{-i2\phi}}, \end{aligned} \quad (3.69)$$

which can be summed to be:

$$t = \frac{t_1 t_3 e^{-i\phi}}{1 + r_1 r_2 e^{-i2\phi}}. \quad (3.70)$$

Equations (3.68) and (3.70) are very useful when studying the optical properties of thin films coated on a substrate as discussed by *Gould* in this handbook. (The reader is referred to the chapter on Thin Films by Gould.) In practice, the two most popular approaches are either to analyze optical transmission spectra, which may be observed using a standard spectrophotometer, or ellipsometric investigations of reflection. Both approaches are briefly explained below.

3.4.1 Swanepoel's Analysis of Optical Transmission Spectra

One of the simplest and most practically realizable approaches to the problem was developed by *Swanepoel* [3.73]. He showed that for a uniform film with a thickness d , a refractive index n and an absorption coefficient α deposited on the substrate with a refractive index s , the transmittance can be expressed as

$$T(\lambda) = \frac{Ax}{B - Cx \cos \varphi + Dx^2}, \quad (3.71a)$$

where

$$A = 16n^2 s, \quad (3.71b)$$

$$B = (n+1)^3 (n+s^2), \quad (3.71c)$$

$$C = 2(n^2 - 1)(n^2 - s^2), \quad (3.71d)$$

$$D = (n-1)^3 (n-s^2), \quad (3.71e)$$

$$\varphi = 4\pi n d / \lambda, \quad (3.71f)$$

$$x = \exp(-\alpha d), \quad (3.71g)$$

and n is a function of λ .

For a nonuniform film with a wedge-like cross-section, (3.71a) must be integrated over the thickness of the film, giving

$$T_{\Delta d}(\lambda) = \frac{1}{\varphi_2 - \varphi_1} \int_{\varphi_1}^{\varphi_2} \frac{Ax}{B - Cx \cos \varphi + Dx^2} d\varphi, \quad (3.72)$$

where

$$\varphi_1 = 4\pi n(d - \Delta d) / \lambda$$

and

$$\varphi_2 = 4\pi n(d + \Delta d) / \lambda, \quad (3.73)$$

where d is the average thickness of the film and Δd is the variation of the thickness throughout the illumination (testing) area. The concept behind Swanepoel's method is to construct two envelopes $T_M(\lambda)$ and $T_m(\lambda)$ that pass through the maxima and minima of $T(\lambda)$ and to split the entire spectral range into three regions: negligible absorption, weak absorption and strong absorption regions, as shown in Fig. 3.16.

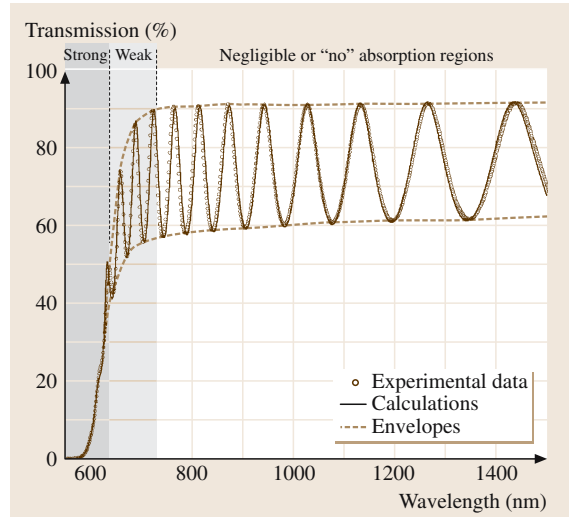


Fig. 3.16 Optical transmission of a-Se thin film. Calculations are done using (3.71) with the $n(\lambda)$ and $\alpha(\lambda)$ relations shown in Fig. 3.17. The film was prepared by the thermal evaporation of photoreceptor-grade selenium pellets. Film thickness was $2 \mu\text{m}$. Tentative regions of strong, weak and negligible absorption are also shown

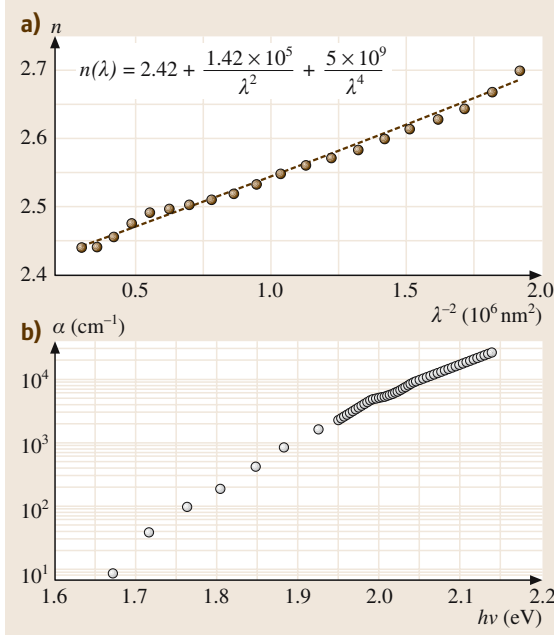


Fig. 3.17 (a) The spectral dependence of the refractive index of the a-Se thin film from Fig. 3.16. The line corresponds to the Cauchy approximation with the parameters shown on the figure. (b) The spectral dependence of the absorption coefficient of the same a-Se thin film

In the region of negligible absorption, (3.72) yields

$$T_{M/m} = \frac{\lambda}{2\pi n \Delta d} \frac{a}{\sqrt{1-b^2}} \tan^{-1} \left[\frac{1 \pm b}{\sqrt{1-b^2}} \tan \left(\frac{2\pi n \Delta d}{\lambda} \right) \right], \quad (3.74)$$

where $a = A/(B+D)$ and $b = C/(B+D)$. The “plus” sign in (3.74) corresponds to transmission maxima and “minus” to minima. Equation (3.74) is used to find Δd and n .

In the region of weak absorption,

$$T_{M/m} = \frac{\lambda}{2\pi n \Delta d} \frac{a_x}{\sqrt{1-b_x^2}} \tan^{-1} \left[\frac{1 \pm b_x}{\sqrt{1-b_x^2}} \tan \left(\frac{2\pi n \Delta d}{\lambda} \right) \right], \quad (3.75)$$

where $a_x = Ax/(B+Dx^2)$ and $b_x = Cx/(B+Dx^2)$. Equation (3.75) allows us to find x and n using the previously found Δd .

The previous results are used to construct $n(\lambda)$ and to create approximations using the Cauchy, Sellmeier or

Wemple–Di Dominico dispersion equations. An example of a Cauchy approximation is shown in Fig. 3.17.

The positions of the extrema are given by the equation

$$2nd = m\lambda, \quad (3.76)$$

where $m = 1, 2, 3, \dots$ for maxima and $m = 1/2, 3/2, \dots$ for minima. Therefore, d can be found from the slope of (n/λ) versus m .

In the region of strong absorption, the absorption coefficient is calculated as

$$x = \frac{A - \sqrt{A^2 - 4T_i^2 BD}}{2T_i D}, \quad (3.77)$$

where

$$T_i = \frac{2T_M T_m}{T_M + T_m} \quad (3.78)$$

and A , B and D are calculated using the above-mentioned $n(\lambda)$ approximation.

It is worth noting that the division into negligible absorption, weak absorption and strong absorption regions is quite arbitrary and should be checked using trial-and-error methods.

3.4.2 Ellipsometry

Ellipsometry measures changes in the polarization of light incident on a sample in order to sensitively characterize surfaces and thin films, for example. Interactions of the polarized light with the sample cause a change in the polarization of the light, which may then be measured by analyzing the light reflected from the sample. The plane-polarized light incident on the sample can be resolved into two components – the component parallel to the plane of incidence (the p-component), and the component perpendicular to it (the s-component). That is, the incident light can be resolved into two components E_p (incident) and E_s (incident), where E refers to the electric field in the light wave, and similarly, the reflected light can be resolved into two components E_p (reflected) and E_s (reflected), enabling so-called Fresnel reflection coefficients, $R_p = E_p$ (reflected)/ E_p (incident) and $R_s = E_s$ (reflected)/ E_s (incident), to be obtained for the p- and s-components, respectively. Ellipsometry measures ρ , the ratio of these two coefficients, which is given by:

$$\rho = (R_p/R_s) = e^{i\Delta} \tan(\Psi), \quad (3.79)$$

where ρ is expressed in terms of the so-called ellipsometric angles Ψ ($0^\circ \leq \Psi \leq 90^\circ$) and Δ ($0^\circ \leq \Delta \leq 360^\circ$). These angles are defined as $\Psi = \tan^{-1} |\rho|$ and the differential phase change, $\Delta = \Delta_p - \Delta_s$. Thus, ellipsometry measures a change in the polarization, expressed as Ψ and Δ , in order to characterize materials. Because ellipsometry measures the ratio of two values, it can be made to be highly accurate and reproducible. The ratio ρ is a complex number; it contains the “phase” information Δ , which makes the measurement very sensitive. However, establishing values for Ψ and Δ is not particularly useful in itself for sample characterization. What one really wants to determine are the parameters of the sample, including, for example, the film thickness, optical constants, and the refractive index. These characteristics can be found by using the measured values of Ψ and Δ in an appropriate model describing the interaction of light with the sample. As an example, consider light reflected off an optically absorbing sample in air (in other words, with a refractive index of unity). The sample can be characterized by a complex refractive index $n - ik$, where n and k are the sample’s refractive index and extinction coefficients at a particular wavelength. From Fresnel’s equations, assuming that the light is incident at an angle ϕ to the sample normal, one gets

$$\begin{aligned} n^2 - k^2 &= \sin^2(\phi) [1 + \{\tan^2(\phi) \\ &\times [\cos^2(2\Psi) - \sin^2(2\Psi) \sin^2(\Delta)] \\ &/ [1 + \sin(2\Psi) \cos(\Delta)]^2\}] \end{aligned} \quad (3.80a)$$

and

$$\begin{aligned} 2nk &= \sin^2(\phi) \tan^2(\phi) \sin(4\Psi) \sin(\Delta) \\ &/ [1 + \sin(2\Psi) \cos(\Delta)]^2. \end{aligned} \quad (3.80b)$$

Since the angle ϕ is set in the experiment, the two parameters measured from the experiment (Ψ and Δ) can be used to deduce the two remaining unknown variables in the equation above – namely, n and k . For a given model (a given set of equations used to describe the sample), the mean squared error between the model and the measured Ψ and Δ values is minimized, typically using the Marquardt–Levenberg algorithm, in order to quickly determine the minimum or best fit within some predetermined confidence limits. Thus, the n and k values are established using this procedure. In practice, ellipsometers consist of a source of linearly polarized light, polarization optics, and a detector. There are a number

of different approaches to conducting an ellipsometry experiment, including null ellipsometers, rotating analyzer/compensator ellipsometers, and spectroscopic ellipsometers. Each of these approaches is discussed briefly below.

Historically, the first ellipsometers that were developed were null ellipsometers. In this configuration, the orientation of the polarizer, compensator and analyzer are adjusted such that the light incident on the detector is extinguished or “nulled”. It should be noted that there are 32 combinations of polarizer, compensator and analyzer angles that can result in a given pair of Ψ and Δ values. However, because any two angles of the polarizer, compensator and analyzer that are 180° apart are optically equivalent, only 16 combinations of angles need to be considered if all angles are restricted to values below 180° . However, even when automated, this approach is thus inherently slow and spectroscopic measurements are difficult to make. However, this configuration can be very accurate and has low systematic errors. In order to speed up measurements, rotating analyzer/polarizer ellipsometers were developed. In these systems, either the analyzer or polarizer is continuously rotated at a constant angular velocity (typically about 100 Hz) about the optical axis. The operating characteristics of both of these configurations are similar. However, the rotating polarizer system requires the light source to be totally unpolarized. Any residual polarization in the source results in a source of measurement error unless corrected. Similarly, a rotating analyzer system is susceptible to the polarization sensitivity of the detector. However, solid state semiconductor photodetectors have extremely high polarization sensitivities. Thus, commercial systems tend to use rotating analyzer systems where residual polarization in the source is not an issue. In this case, the (sinusoidal) variation in the amplitude of the detector signal can be directly related to the ellipticity of the reflected light – Fourier analysis of this output provides values for Ψ and Δ . Such systems can provide high-speed, accurate measurements and the lack of a compensator actually improves the measurement by eliminating any errors associated with these components. Spectroscopic ellipsometers extend the concepts developed for measurement at a single wavelength to measurements at multiple wavelengths – typically as many as 40 wavelengths. Being able to measure the dispersion in optical constants with wavelength adds another dimension to the analysis, permitting unambiguous determinations of material and structure parameters.

3.5 Optical Materials

3.5.1 Abbe Number or Constringence

In an optical medium, the Abbe number is defined as the inverse of its dispersive power; that is, it represents the relative importance of refraction and dispersion. There are two common definitions based on the use of different standard wavelengths. The Abbe number ν_d is defined by:

$$\nu_d = (n_d - 1)/(n_F - n_C), \tag{3.81}$$

where n_F , n_d and n_C are the refractive indices of the medium at the Fraunhofer standard wavelengths corresponding to the helium d-line ($\lambda_d = 587.6$ nm, yellow), the hydrogen F-line ($\lambda_F = 486.1$ nm, blue) and the hy-

drogen C-line ($\lambda_C = 656.3$ nm, red), respectively. The Abbe number of a few glasses are listed in Table 3.10. The Abbe number ν_e , on the other hand, is defined by

$$\nu_e = (n_e - 1)/(n_{F'} - n_{C'}) : \tag{3.82}$$

where n_e , $n_{F'}$ and $n_{C'}$ are the refractive indices at wavelengths of the e-line (546.07 nm), the F'-line (479.99 nm) and the C'-line (643.85 nm), respectively.

The relationship between the refractive index n_d and the Abbe number ν_d is given by the Abbe diagram, where n_d is plotted on the y-axis against its corresponding ν_d value for different glasses, usually with the Abbe numbers decreasing on the x-axis, as shown in Fig. 3.18.

3.5.2 Optical Materials

Optical materials are those crystalline solids and glasses that are commonly used in the construction of optical components, such as lenses, prisms, window, mirrors and polarizers. In addition to having the required optical properties, such as a well defined refractive index n with a known dependence on wavelength $n = n(\lambda)$ and temperature, optical materials should also have various other desirable material properties, such as good homogeneity (including negligible macroscopic variations in the refractive index), negligible thermal expansion (small α_L , thermal expansion coefficient), resistance to mechanical damage (resistance to scratching), and resistance to chemical degradation (chemical corrosion, staining). Also, in the case of optical glasses, it is desirable to have negligible air (or gas) bubbles incorporated into the glass structure during fabrication (during shaping for example), negligible stress-induced birefringence, negligible nonlinear properties (unless used specifically due to its nonlinearity), and negligible fluorescence. In all of these cases, negligible implies less than a tolerable quantity within the context of the particular optical application. For example, at a given wavelength and for many glass materials used in optics, n should not vary by more than 10^{-5} , while 10^{-6} is required for certain optics applications in astronomy. In addition, they should have good or reliable man-

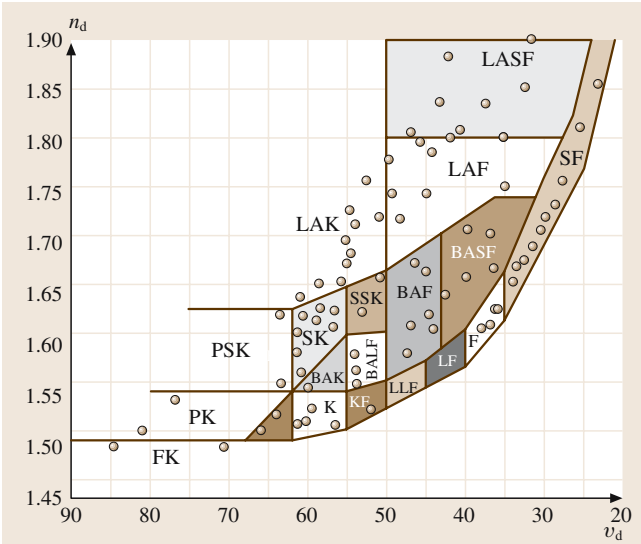


Fig. 3.18 The Abbe diagram is a diagram in which the refractive indices n_d of glasses are plotted against their Abbe numbers in a linear n_d versus ν_d plot and, usually, with the Abbe number decreasing along the x-axis, rather than increasing. A last letter of F or K represents flint or crown glass. Other symbols are as follows: S, dense, L, light; LL, extra light; B, borosilicate; P, phosphate; LA or La, lanthanum; BA or Ba, barium. Examples: BK, dense flint; LF, light flint; LLF, extra light flint; SSK, extra dense crown; PK, phosphate crown; BAK, barium crown; LAF, lanthanum flint

Table 3.10 Abbe numbers for a few glasses. PC denotes polycarbonate, PMMA is polymethylmethacrylate, and PS represents polystyrene

Optical glass →	SF11	F2	BaK1	Crown glass	Fused silica	PC	PMMA	PS
ν_d →	25.76	36.37	57.55	58.55	67.80	34	57	31

Table 3.11 The refractive indices, n_d , and Abbe numbers, v_d , (3.81) of selected optical materials (compiled from the websites of Oriel, Newport and Melles-Griot); n_d at $\lambda_d = 587.6$ nm, α_L is the linear thermal expansion coefficient

Glasses	Transmission (typical, nm)	n_d	v_d	Applications	Comment
Fused silica	175–2000	1.45846	67.8	Lenses, windows, prisms, interferometric FT-IR components. UV lithography	Synthetic. Has UV properties; transmittance and excellent thermal low α_L . Resistant to scratching
SF 11, flint	380–2350	1.78472	25.76	Lenses, prisms	Flint glasses have $v_d < 50$
LaSFN9, flint	420–2300	1.85025	32.17	Lenses, prisms	High refractive index. More lens power for less curvature
BK7, borosilicate crown	380–2100	1.51680	64.17	Visible and near IR optics. Lenses, windows, prisms, interferometric components	All around excellent optical lens material. Not recommended for temperature-sensitive applications
BaK1, barium crown	380–2100	1.57250	57.55	Visible and near IR optics. Lenses, windows, prisms, interferometric components	All around excellent optical lens material. Not recommended for temperature-sensitive applications
Optical crown	380–2100	1.52288	58.5	Lenses, windows, prisms, interferometric components	Lower quality than BK7
Pyrex, borosilicate glass		1.43385	66	Mirrors	Low thermal expansion
Crystals					
CaF ₂ crystal	170–7000	1.43385	94.96	Lenses, windows for UV optics, especially for excimer laser optics	Sensitive to thermal shock
MgF ₂ crystal	150–7000	$n_0 = 1.37774$ $n_e = 1.38956$		Lenses, windows, polarizers, UV transmittance	Positive birefringent crystal. Resistant to thermal and mechanical shock
Quartz, SiO ₂ crystal	150–2500	$n_0 = 1.54431$ $n_e = 1.55343$		UV optics. Wave plates. Polarizers	Positive uniaxial birefringent crystal
Sapphire, Al ₂ O ₃ crystal	150–6000	1.7708 (546.1 nm)		UV-Far IR windows, high power laser optics	High surface hardness, Scratch resistant. Chemically inert
Auxiliary optical materials					
ULE SiO ₂ -TiO ₂ glass				Optical spacers	Very small thermal expansion
Zerodur, glass ceramic composite		1.5424	56–66	Mirror substrates. Not suitable for transmission optics due to internal scattering	Ultra-low α_L . Fine mixture of glass and ceramic crystals (very small size)

ufacturability at an affordable cost. There are various useful optical materials which encompass not only single crystals (such as CaF_2 , MgF_2 , quartz, sapphire) but also a vast range of glasses (which are supercooled liquids with high viscosity, such as flint and crown glasses as well as fused silica). Higher refractive index materials have more refractive power and allow lens designs that need less curvature to focus light, and hence tend to give fewer aberrations. Flint glasses have a larger refractive index than crown glasses. On the other hand, crown glasses are chemically more stable, and can be produced more to specification. While most optical materials are used for their optical properties (such as in optical transmission), certain “optical” materials (auxiliary materials) are used in optical applications such as mirror substrates and optical spacers for their nonoptical properties, such as their negligible thermal expansion coefficients. Some optical properties of selected optical materials and their applications are listed in Table 3.11.

3.5.3 Optical Glasses

Optical glasses are a range of noncrystalline transparent solids used to fabricate various optical components, such as lenses, prisms, light pipes and windows. Most (but not all) optical glasses are either crown (K) types or flint (F) types. K-glasses are usually soda-lime-silica glasses, whereas flint glasses contain substantial lead oxide; hence F-glasses are denser and have higher refractive powers and dispersions. Barium glasses contain barium oxide instead of lead oxide and, like lead glasses, have high refractive indices, but lower dispersions. There are

other high refractive index glasses, such as lanthanum- and rare earth-containing glasses. Optical glasses can also be made from various other glass formers, such as boron oxide, phosphorus oxide and germanium oxide. The Schott glass code or number is a special number designation (511 604.253 for Schott glass K7) in which the first three numbers (511) represent the three decimal places in the refractive index ($n_d = 1.511$), the next three numbers (604) represent the Abbe number ($v_d = 60.4$), and the three numbers after the decimal (253) represent the density ($\rho = 2.53 \text{ g/cm}^3$). A different numbering system is also used, where a colon is used to separate n_d and v_d ; for example, 517:645 for a particular borosilicate crown means $n_d = 1.517$, $v_d = 64.5$ (see also Sect. 3.5.1).

In the Schott glass coding system, optical glasses are represented by letters in which a last letter of K refers to crown, and F to flint. The first letters usually represent the most important component in the glass, such as P in the case of phosphate. The letters Kz (“Kurtz”), L (“leicht”) and S (“schwer”) before K or F represent short, light and dense (heavy) respectively (from German). S after K or F means “special”. Examples include: BK, borosilicate crown; FK, fluor crown; PK, phosphate crown; PSK, dense phosphate crown; BaLK, light barium crown; BaK, barium crown; BaSK, dense barium crown; SSK, extra dense barium crown; ZnK, zinc crown; LaK, lanthanum crown; LaSK, dense lanthanum crown; KF, crown flint; SF, dense flint; SFS, special dense flint; BaF, barium flint; BaLF, barium light flint; BaSF, dense barium flint; LLF, extra light flint; LaF, lanthanum flint.

References

- 3.1 S. Adachi: *Properties of Group IV, III–V and II–VI Semiconductors* (Wiley, Chichester, UK 2005)
- 3.2 M. Fox: *Optical Properties of Solids* (Oxford Univ. Press, Oxford 2001)
- 3.3 J. H. Simmons, K. S. Potter: *Optical Materials* (Academic, San Diego 2000)
- 3.4 D. E. Aspnes, A. A. Studna: *Phys. Rev.* **B27**, 985 (1983)
- 3.5 A. L. Cauchy: *Bull. Sci. Math.* **14**, 6 (1830)
- 3.6 A. L. Cauchy: *Mémoire sur la Dispersion de la Lumière* (Calve, Prague 1836)
- 3.7 D. Y. Smith, M. Inokuti, W. Karstens: *J. Phys.: Condens. Mat.* **13**, 3883 (2001)
- 3.8 D. F. Edwards, E. Ochoa: *Appl. Opt.* **19**, 4130 (1980)
- 3.9 W. L. Wolfe: *The Handbook of Optics*, ed. by W. G. Driscoll, W. Vaughan (McGraw-Hill, New York 1978)
- 3.10 J. W. Fleming: *Appl. Opt.* **23**, 4486 (1984)
- 3.11 K. L. Wolf, K. F. Herzfeld, H. Geiger, K. Scheel (eds.): *Handbuch der Physik*, Vol. 20 (Springer, Berlin, Heidelberg 1928)
- 3.12 M. Herzberger: *Opt. Acta* **6**, 197 (1959)
- 3.13 P. J. Herve, L. K. J. Vandamme: *J. Appl. Phys.* **77**, 5476 (1995)
- 3.14 D. Dale, F. Gladstone: *Philos. Trans.* **153**, 317 (1863)
- 3.15 S. H. Wemple, M. DiDomenico Jr.: *Phys. Rev.* **3**, 1338 (1971)
- 3.16 W. J. Turner, W. E. Reese: *Phys. Rev.* **127**, 126 (1962)
- 3.17 W. G. Spitzer, H. Y. Fan: *Phys. Rev.* **106**, 882 (1957)
- 3.18 J. D. Wiley, M. DiDomenico: *Phys. Rev.* **B1**, 1655 (1970)
- 3.19 H. R. Riedl: *Phys. Rev.* **127**, 162 (1962)
- 3.20 R. L. Weihler: *Phys. Rev.* **152**, 735 (1966)
- 3.21 E. Hagen, H. Rubens: *Ann. Phys.* **14**, 986 (1904)

- 3.22 R. J. Elliott, A. F. Gibson: *An Introduction to Solid State Physics and Its Applications* (Macmillan, London 1974)
- 3.23 H. B. Briggs, R. C. Fletcher: *Phys. Rev.* **91**, 1342 (1953)
- 3.24 C. R. Pidgeon: Optical Properties of Solids. In: *Handbook on Semiconductors*, Vol. 2, ed. by M. Balkanski (North Holland, Amsterdam 1980) Chap. 5, pp. 223–328
- 3.25 H. E. Ruda: *J. Appl. Phys.* **72**, 1648 (1992)
- 3.26 H. E. Ruda: *J. Appl. Phys.* **61**, 3035 (1987)
- 3.27 W. Kaiser, R. J. Collins, H. Y. Fan: *Phys. Rev.* **91**, 1380 (1953)
- 3.28 I. Kudman, T. Seidel: *J. Appl. Phys.* **33**, 771 (1962)
- 3.29 A. E. Rakhshani: *J. Appl. Phys.* **81**, 7988 (1997)
- 3.30 R. H. Bube: *Electronic Properties of Crystalline Solids* (Academic, San Diego 1974)
- 3.31 F. Urbach: *Phys. Rev.* **92**, 1324 (1953)
- 3.32 J. Pankove: *Phys. Rev.* **140**, 2059 (1965)
- 3.33 R. A. Street: *Hydrogenated Amorphous Silicon* (Cambridge Univ. Press, Cambridge 1991)
- 3.34 D. A. Papaconstantopoulos, E. N. Economou: *Phys. Rev.* **24**, 7233 (1981)
- 3.35 M. H. Cohen, Fritzsche, S. R. Ovshinsky: *Phys. Rev. Lett.* **22**, 1065 (1969)
- 3.36 N. F. Mott, E. A. Davis: *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford 1979)
- 3.37 J. Singh, K. Shimakawa: *Advances in Amorphous Semiconductors* (Taylor & Francis, London 2003)
- 3.38 S. Sherman, S. Wagner, R. A. Gottscho: *Appl. Phys. Lett.* **69**, 3242 (1996)
- 3.39 T. Tiedje, J. M. Cebulla, D. L. Morel, B. Abeles: *Phys. Rev. Lett.* **46**, 1425 (1981)
- 3.40 K. Winer, L. Ley: *Phys. Rev.* **36**, 6072 (1987)
- 3.41 D. P. Webb, X. C. Zou, Y. C. Chan, Y. W. Lam, S. H. Lin, X. Y. Lin, K. X. Lin, S. K. O'Leary, P. K. Lim: *Sol. State Commun.* **105**, 239 (1998)
- 3.42 W. B. Jackson, S. M. Kelso, C. C. Tsai, J. W. Allen, S.-H. Oh: *Phys. Rev.* **31**, 5187 (1985)
- 3.43 S. K. O'Leary, S. R. Johnson, P. K. Lim: *J. Appl. Phys.* **82**, 3334 (1997)
- 3.44 S. M. Malik, S. K. O'Leary: *J. Non-Cryst. Sol.* **336**, 64 (2004)
- 3.45 S. R. Elliott: *The Physics and Chemistry of Solids* (Wiley, Sussex 1998)
- 3.46 J. Singh: *Nonlin. Opt.* **29**, 119 (2002)
- 3.47 J. Tauc: *Phys. Stat. Solidi* **15**, 627 (1966)
- 3.48 F. Orapunt, S. K. O'Leary: *Appl. Phys. Lett.* **84**, 523 (2004)
- 3.49 A. P. Sokolov, A. P. Shebanin, O. A. Golikova, M. M. Mezdrogina: *J. Phys. Condens. Mat.* **3**, 9887 (1991)
- 3.50 G. D. Cody: *Semicond. Semimet.* **21**, 11 (1984)
- 3.51 K. Morigaki: *Physics of Amorphous Semiconductors* (World Scientific, London 1999)
- 3.52 L. Ley: *The Physics of Hydrogenated Amorphous Silicon II*, ed. by J. D. Joannopoulos, G. Lukovsky (Springer, Berlin, Heidelberg 1984) p. 61
- 3.53 T. Aoki, H. Shimada, N. Hirao, N. Yoshida, K. Shimakawa, S. R. Elliott: *Phys. Rev.* **59**, 1579 (1999)
- 3.54 K. Shimakawa, Y. Ikeda, S. Kugler: *Non-Crystalline Materials for Optoelectronics*, Optoelectronic Materials and Devices, Vol. 1 (INOE Publ., Bucharest 2004) Chap. 5, pp. 103–130
- 3.55 J. Singh: *J. Mater. Sci. Mater. El.* **14**, 171 (2003)
- 3.56 S. M. Malik, S. K. O'Leary: *J. Mater. Sci. Mater. El.* **16**, 177 (2005)
- 3.57 S. Abe, Y. Toyozawa: *J. Phys. Soc. Jpn.* **50**, 2185 (1981)
- 3.58 A. O. Kodolbas: *Mater. Sci. Eng.* **98**, 161 (2003)
- 3.59 J. Singh: *Excitation Energy Transfer Processes in Condensed Matter* (Plenum, New York 1994)
- 3.60 J. Singh: *Nonlin. Opt.* **18**, 171 (1997)
- 3.61 J. Singh, T. Aoki, K. Shimakawa: *Philos. Mag.* **82**, 855 (2002)
- 3.62 R. J. Elliott: *Polarons and Excitons*, ed. by K. G. Kuiper, G. D. Whitfield (Oliver Boyd, London 1962) p. 269
- 3.63 T. Aoki, S. Komodoori, S. Kobayashi, T. Shimizu, A. Ganjoo, K. Shimakawa: *Nonlin. Opt.* **29**, 273 (2002)
- 3.64 P. C. Becker, N. A. Olsson, J. R. Simpson: *Erbium-Doped Fiber Amplifiers. Fundamentals and Technology* (Academic, New York 1999)
- 3.65 B. Henderson, G. F. Imbusch: *Optical Spectroscopy of Inorganic Solids* (Clarendon, Oxford 1989)
- 3.66 S. Hufner: *Optical Spectra of Rare Earth Compounds* (Academic, New York 1978)
- 3.67 G. Blasse, B. C. Grabmaier: *Luminescent Materials* (Springer, Berlin, Heidelberg 1994)
- 3.68 B. DiBartolo: *Optical Interactions in Solids* (Wiley, New York 1968)
- 3.69 W. T. Carnall, G. L. Goodman, K. Rajnak, R. S. Rana: *J. Chem. Phys.* **90**, 3443 (1989)
- 3.70 M. J. Weber: *Phys. Rev.* **157**, 262 (1967)
- 3.71 E. Desurvire: *Erbium-Doped Fibre Amplifiers* (Wiley, New York 1994)
- 3.72 K. A. Gschneidner, Jr., LeRoy, Eyring (Eds.): *Handbook on the Physics and Chemistry of Rare Earths*, Vol. 25 (Elsevier, Amsterdam 1998)
- 3.73 R. Swanepoel: *J. Phys. E* **17**, 896 (1984)

4. Magnetic Properties of Electronic Materials

This work reviews basic concepts from both traditional macroscopic magnetism and unconventional magnetism, in order to understand current and future developments of submicronic spin-based electronics, where the interplay of electronic and magnetic properties is crucial. Traditional magnetism is based on macroscopic observation and physical quantities are deduced from classical electromagnetism. Physical interpretations are usually made with reference to atomic magnetism, where localized magnetic moments and atomic physics prevail, despite the fact that standard ferromagnetic materials such as Fe, Co and Ni are not localized-type magnets (they have extended s and localised d electronic states). While this picture might be enough to understand some aspects of traditional storage and electromechanics, it is not sufficient when describing condensed matter systems with smaller length scales (progressing toward the nanometer range). In this case, the precise nature of the magnetism (localized, free or itinerant as in Fe, Co and Ni transition metals) should be accounted for, along with the simultaneous presence of charge and spin on carriers. In addition, when we deal with the thin films or multilayers found in conventional electronics, or with objects of reduced dimensionality (such as wires, pillars, dots or grains), the magnetic properties are expected to be different from conventional three-dimensional bulk systems.

This chapter is organized as follows. We begin (in the Introduction) by highlighting the new era of submicronic spin-based electronics, and we

4.1	Traditional Magnetism	81
4.1.1	Fundamental Magnetic Quantities	81
4.1.2	The Hysteresis Loop	83
4.1.3	Intrinsic Magnetic Properties	87
4.1.4	Traditional Types of Magnetism and Classes of Magnetic Materials	90
4.2	Unconventional Magnetism	93
4.2.1	Conventional and Unconventional Types of Exchange and Coupling in Magnetic Materials	93
4.2.2	Engineering and Growth of Thin Magnetic Films	94
4.2.3	Electronic Properties: Localized, Free and Itinerant Magnetism and Spin-Polarised Band Structure	95
4.2.4	Prospects for Spintronics and Quantum Information Devices	98
	References	99

present a table of papers on the topics we cover in the chapter, for the reader who wishes to learn more. The traditional elements of magnetism, such as the hysteresis loop, conventional types of magnetism and magnetic materials, are then presented (in Sect. 4.1). We then briefly describe (in Sect. 4.2) unconventional magnetism, which can be used to understand new high-tech materials that will be used in future devices based on spintronics and quantum information.

Digital information technology involves three main activities:

- Processing of information (using transistors, logic gates, CPU, RAM, DSP...)
- Communication of information (using networks, switches, cables, fibers, antennae...)

- Storage of information (using tapes, hard disks, CD, DVD, flash memory...)

The application of magnetism to such technologies has traditionally been confined to information storage, originating from the development of bubble and ferrite core technologies, when RAM memory was based on

Table 4.1 Selected topics in magnetism, with corresponding applications and references for further reading

Topic	Applications/comments	Reference
Amorphous magnets	Shielding, sensing, transformers, transducers	<i>Boll, Warlimont</i> [4.1]
Coherent rotation model	Hysteresis loop determination	<i>Stoner, Wohlfarth</i> [4.2]
Coupling and exchange in multilayers	Biquadratic exchange, exchange bias, spring magnets	<i>Platt et al.</i> [4.3], <i>Slonczewski</i> [4.4], <i>Koon</i> [4.5]
Giant magnetoimpedance	Sensing, detectors	<i>Tannous, Gieraltowski</i> [4.6]
Giant magnetoresistance (<i>GMR</i>)	Recording heads, spin valves, spin filters	<i>White</i> [4.7]
Giant magnetostriction	Smart plane wings, MEMS, actuators, transducers, resonators	<i>Schatz et al.</i> [4.8], <i>Dapino et al.</i> [4.9]
Itinerant magnetism	Magnetism in transition metals	<i>Himpsel et al.</i> [4.10]
Localized magnetism	Atoms/molecules/ions/insulators	<i>Jansen</i> [4.11]
Losses in magnetic materials	Eddy currents, hysteresis loss	<i>Goodenough</i> [4.12]
Magnetic recording	Hard disk technology	<i>Richter</i> [4.13]
Magnetic thin films	Growth and characterization	<i>Himpsel et al.</i> [4.10]
Magnetoelastic effects	Cantilevers, MEMS	<i>Farber et al.</i> [4.14], <i>Dapino et al.</i> [4.9]
Microwave devices	Communications, bubble memory	<i>Coeure</i> [4.15], <i>Wigen</i> [4.16]
Permanent or hard magnets	Relays, motors, transformers	<i>Gutfleisch</i> [4.17]
Quantum computing/communications	Quantum devices, magnetic RAM	<i>Burkard, Loss</i> [4.18]
Sensors	Field detectors, probes	<i>Hauser et al.</i> [4.19]
Soft magnets	Shielding, sensing, transformers, transducers	<i>Boll, Warlimont</i> [4.1]
Spintronics	Spin diode, spin <i>LED</i> , spin transistor, magnetic RAM	<i>Prinz</i> [4.20], <i>Zutic et al.</i> [4.21]
Technology overview	Permanent and soft magnets	<i>Kronmueller</i> [4.22], <i>Simonds</i> [4.23]
Types of magnetic order	Ferromagnetism, antiferromagnetism, diamagnetism, paramagnetism	<i>Hurd</i> [4.24]

magnetism. RAM memory is a special type of nonpermanent (primary) information storage device, which can be distinguished from permanent (secondary) or mass information storage devices such as tapes, hard disks, floppy and zip disks, CDs, *DVDs*, and flash memory.

The field of applied magnetism is currently undergoing much transformation due to several recent developments, among which is the progress toward the nanometer scale in integrated circuits. At this length scale, quantum effects become extremely important and carrier spin becomes of interest since it may be conserved over this length scale and so could be used to carry and manipulate information. This would pave the way towards the fabrication of new devices based on charge and spin (spintronic devices) instead of just charge, as used in traditional microelectronics. This means that new types of junctions and transistors could be built that would use magnetism to tackle the processing of information.

Quantum effects have already been used in many products, since they are the basis of the *GMR* effect (Table 4.1) that is the basis for the recent surge towards

extremely high densities in hard disks; however, the longer-term intention is to make use of these effects in basic components of *quantum computers* and *quantum communication devices*. The quantum computer is based on the *qubit* (quantum bit), which is the basic unit of information used in a quantum computer (equivalent to the classical bit used in conventional computing). If we consider a sphere, the classical bit can be viewed as an object with two possible states at the north and the south poles of the sphere, whereas a qubit is an object that can sit anywhere on the surface of the sphere (called the *Bloch sphere*; see *Nielsen and Chuang* [4.25]). A quantum computer can perform a massive number of computations in parallel, since one is allowed to access any superposition of states at any time in quantum mechanics due to its linearity, in contrast to a classical von Neumann type of computer, where one only has access to a single state at any time (Table 4.1). Quantum communications are extremely secure, since a caller may build a coherent state with the callee where any intrusion can perturb the coherence, providing very efficient detection.

From a fundamental point of view, one can start by representing a magnetic material by a single magnetic moment and then studying its behavior, before investigating the many interacting moments that are the building blocks of magnetic materials. From an applied point of view, the orientation of the moment defines the value

of the bit. Once the orientation of the moment has been linked to a bit value, it becomes important to understand the physical processes, the energetics and the dynamics of the change in moment orientation (moment reversal from left to right or moment flip from up to down) in order to be able to control, alter and predict the bit value.

4.1 Traditional Magnetism

Classical magnetism relates to magnetic moments and their behavior when an external field, mechanical stress or some other perturbing effect is applied. The idea is to investigate the way that the energy of the magnetic moment changes with time. In order to describe the different magnetic energy terms that control the behavior of a moment, we can start from a single isolated moment in vacuum, at zero temperature ($T = 0$ K), zero applied magnetic field ($\mathbf{H} = 0$) and zero frequency ($f = 0$). It is obvious that the energy is zero from a classical point of view (quantum mechanically, however, it is worth noting that, even at $T = 0$ K, quantum fluctuations exist that could disorient, flip or reverse the moment).

The next step is to apply a magnetic field \mathbf{H} ($T = 0$ K, $f = 0$); we then obtain the Zeeman energy $E_Z = -\mathbf{M} \cdot \mathbf{H}$, with \mathbf{M} representing the moment. If we place the moment in an unbounded (of infinite size) crystal, it is clear that the energy of the moment is anisotropic, since the direction of the moment (called an easy axis direction) is imposed by the internal symmetry of the crystal, in contrast to the case in vacuum, where \mathbf{M} is free to point in any direction. The crystal may possess a set of easy axes (easy planes), in which case the moment will point in one of several directions (or any direction in one of several planes).

If the body containing the moment has a finite size, a new energy must be accounted for: the demagnetization energy. Magnetic surface charges (poles) induced on the surface bounding the body create a demagnetizing field inside the body (this is called the stray field outside of the body). The demagnetization energy is also called the shape anisotropy energy or the magnetostatic energy because it is (approximately) expressed (for ellipsoidal bodies) as $2\pi N_{\alpha\beta} \mathbf{M}_\alpha \mathbf{M}_\beta$, where $N_{\alpha\beta}$ is a set of factors (demagnetization coefficients) that depend on the shape of the body (the Einstein summation convention is used for repeated indices).

Now suppose we include several local moments in a material. If sites i and j carrying moments \mathbf{M}_i and \mathbf{M}_j

are close enough, we then have an interaction energy between them called the exchange energy $-A_{ij} \mathbf{M}_i \mathbf{M}_j$, which will align moments \mathbf{M}_i and \mathbf{M}_j if A_{ij} is positive (ferromagnetic interaction). If A_{ij} is negative, the moments will align antiparallel to each other (antiferromagnetic interaction). In a ferromagnet we have a net moment, whereas in an antiferromagnet the net moment is zero.

The above description considers localized magnetism in distinct atoms (such as in a gas), ions, molecules or in special materials (like insulators or rare-earth solid state compounds possessing external-shell f electrons with highly atomic-like character). If we have a conducting material with free electrons interacting with localized atoms/ions/molecules (for transition metals with s and d electrons or magnetic semiconductors for example), a different type of magnetism called itinerant magnetism occurs. Nevertheless, it is possible to extend the notion of the magnetic moment to this case, accounting for the combined effects of free and localized charges modeled as an effective number of Bohr magnetons (see Sect. 4.2.1, Table 4.1 and Table 4.2).

The different physical mechanisms and types of magnetism briefly described above operate at different length scales. In order to gain some perspective and be able to ponder what lies ahead in terms of possible developments and hurdles, Fig. 4.1 gives a summary of different mechanisms, characterization methods and manufacturing processes along with their corresponding length scales. Note that, on the nanometer scale, the device size becomes comparable to most ranges of interactions encountered in magnetic materials, and this will trigger the development of novel effects and devices.

4.1.1 Fundamental Magnetic Quantities

Magnetization is the fundamental property exhibited by a magnetic substance. It originates from its electrons, as with the electric dipole moment. It can be

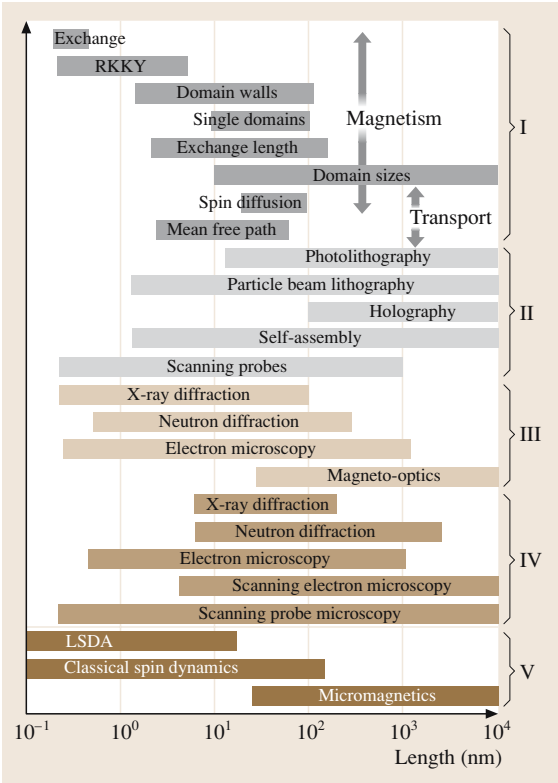


Fig. 4.1 Typical lengths of interest in magnetic materials and spintronics (in relation to basic phenomena, interactions and transport). Diagnostic techniques in III are depth profiling techniques and IV is for surface probing techniques) of interest in magnetic materials are also shown along with their probing lengths. II indicates growth methods and V theoretical characterization techniques. **LSDA** (local spin density approximation) is the spin counterpart of the density function theory (**DFT**) used to calculate band structure. Since most interaction lengths (shown in I) range between 1 and 100 nm, device size is comparable to interaction range in nanoelectronics (after [4.28])

intrinsic (it can exist without the application of any external field) or it can be induced by an external magnetic field. In atoms/molecules/ions and rare-earth solids, magnetization is created from individual atom-like localized magnetic moments. In the solid state, the moment is generally defined by the band structure and an effective moment can be defined in terms of a *Bohr magneton* $\mu_B = e\hbar/2m$, the magnetic moment carried by a single electron (e and m are the electron charge and mass). Magnetization is a thermodynamic quantity that changes with temperature, mechanical

Table 4.2 Selected ferromagnetic solids with their saturation magnetizations M_s , effective Bohr magnetons n_B and Curie temperatures [4.27]

Substance	M_s (G) at 300 K	M_s (G) at 0 K	n_B at 0 K	Curie T_c (K)
Fe	1707	1740	2.22	1043
Co	1400	1446	1.72	1388
Ni	485	510	0.62	627
Gd	–	2060	7.63	292
Dy	–	2920	10.2	88
MnAs	670	870	3.4	318
MnBi	620	680	3.52	630
CrO ₂	515	–	2.03	386
EuO	–	1920	6.8	69

stress and chemical processes such as alloying (see *Chikazumi* [4.26]).

When a field is applied to a magnetic substance, the largest acquired magnetization M_s measured along the direction of the applied field is the saturation magnetization, meaning that all moments are aligned parallel to the field. Since increasing the temperature T causes more moments to misalign, one can define the saturation at $T = 0$ K with $M_s = Nn_B\mu_B$, where N is the number of ions, atoms or molecules, and n_B the effective number of Bohr magnetons (see Table 4.2).

The number n_B is different for atoms/ions/molecules and solids and is determined by the electronic state; for example, for Fe^{3+} ions it is 5 (according to Hund rules, see *Jansen* [4.11]), since we have a $3d^5$ configuration (the orbital contribution is neglected), whereas for Fe atoms it is 2 ($3d^8$ configuration). Trivalent rare-earth ions with highly localized 4f orbitals possess magnetic moments that are determined by $g_J[J(J+1)]^{1/2}$, where g_J is the Landé factor and J the total angular momentum. Transition metal ions possess magnetic moments that are determined by $2[S(S+1)]^{1/2}$, where S is the total spin of the ion (there is no orbital contribution because of orbital moment quenching, see *Kittel* [4.27]). In the solid state, for example for ferromagnetic metals, n_B is determined from the band structure (Table 4.1). At lower dimensions (clusters, dots, thin films...), n_B tends to increase because of lower symmetry and coordination (high symmetry and coordination tend to decrease it). Among the elements, the rare-earth ions (Dy^{3+} , Ho^{3+} and Er^{3+}) possess large values of n_B – on the order of 10 – whereas solids such as Gd-based garnets (see *Wigen* [4.16]) have n_B values on the order of 15 (at $T = 0$ K).

The moment due to angular momentum is given in the localized case (atoms/ions/molecules) by: $-g_L\mu_B\mathbf{L}$ (for orbital angular momentum \mathbf{L}), where $g_L = 1$ (the orbital Landé factor), or by $-g_S\mu_B\mathbf{S}$ (for spin angular momentum \mathbf{S}), where $g_S = 2$ (the spin Landé factor). In the case of an atom/ion with a total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$, the moment is given by $-g_J\mu_B\mathbf{J}$, where:

$$g_J = 1 + [J(J+1) + S(S+1) - L(L+1)]/2J(J+1)$$

which is the total Landé factor. In the solid state, the Landé factor is determined from the band structure (using the spin-splitting of electron or hole bands in the presence of a magnetic field). It may become negative, anisotropic (a tensor) and depend on the wavevector \mathbf{k} . Typical values of g for semiconductors at $\mathbf{k} = 0$ include: InSb, $g = -44$; GaAs, $g = 0.32$; InAs, $g = -12$ [4.11, 21].

4.1.2 The Hysteresis Loop

When a magnetic field is applied to a ferromagnetic material [4.27], a change in magnetization takes place. The straightforward way to understand the underlying phenomena is to plot the magnetization \mathbf{M} along the direction of the applied field \mathbf{H} . The locus of the magnetization depicted in the \mathbf{M} - \mathbf{H} plane is the hysteresis loop (Fig. 4.2).

The term hysteresis means that when the material is field-cycled (in other words, when the field \mathbf{H} is increased then decreased), two different nonoverlapping curves $\mathbf{M}(\mathbf{H})$ are obtained. Another approach involves plotting the induction \mathbf{B} versus \mathbf{H} , since it is \mathbf{B} that is ac-

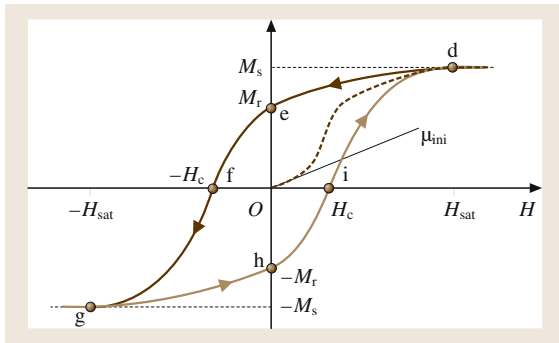


Fig. 4.2 Hysteresis loop showing the fundamental points in the \mathbf{M} - \mathbf{H} plane, such as the saturation magnetization M_s , the remanent M_r , the coercive field H_c (for which $M = 0$) and the saturation field H_{sat} (for which $M = M_s$), as well as the initial magnetization curve (dashed) with initial permeability μ_{ini} (after [4.29])

tually measured. Several points in the \mathbf{M} - \mathbf{H} plane shown in the figure are worth mentioning. Point d is where the magnetization reaches its largest value M_s (the saturation magnetization). Points i and f denote the magnetic fields $\pm H_c$ for which the magnetization is zero. The magnetic field H_c is called the coercive field; it is generally large for hard materials and small for soft magnets. Points e and h denote the remanent magnetization M_r , which is the magnetization that remains after the applied field is switched off, which is the hallmark of permanent magnets.

The ratio M_r/M_s , called the squareness, is close to 1 when the applied magnetic field is closely aligned with the easy axis (and so the hysteresis loop is squarer). Once the easy axis (or set of easy axes) is determined, the angle between the easy axis and the magnetic field (ϕ , say) is varied and the hysteresis loop is plotted for different angles. When the angle ϕ is increased, the opening of the hysteresis loop is reduced; it is largest when the magnetic field is parallel to the easy axis and smallest when the magnetic field is parallel to the so-called hard

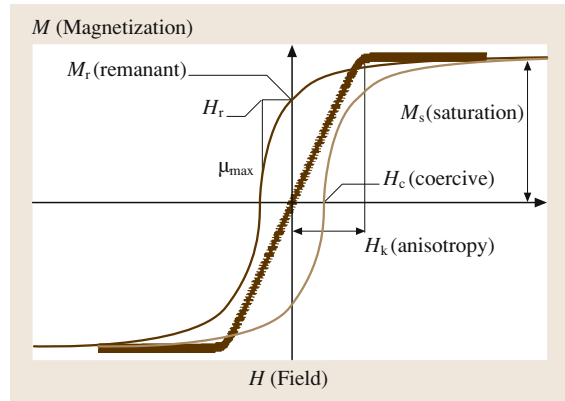


Fig. 4.3 Hysteresis loop obtained for an arbitrary angle ϕ between the magnetic field and the anisotropy axis (easy axis). Associated quantities such as coercive field H_c , remanent magnetization M_r and field H_r (given by the intersection of the tangent to the loop at $-H_c$ and the M_r horizontal line) are shown. The tangent to the loop at $-H_c$ is also called the maximum differential permeability. The *thick line* is the hysteresis loop when the field is along the hard axis and H_k (the anisotropy field) is the field value at the slope break where the magnetization reaches its saturation value M_s . Quantities such as H_c , M_r , H_r and μ_{max} depend on the angle ϕ . When the field is along the easy axis, the coercive field reaches its largest value (the hysteresis loop is broadest), at which point the magnetization jumps (at $\pm H_c$)

axis. All of these characteristics of the hysteresis loop are depicted in Fig. 4.3, and for a given temperature and frequency of the applied field H , quantities such as the remanent magnetization M_r , the remanent field H_r , the coercive field H_c and the maximum differential permeability μ_{\max} vary with the angle ϕ ;

In addition, the hysteresis loop changes and may even disappear altogether above a given temperature (Curie temperature), and it can be seriously altered by changing the frequency of the field (see below).

The Magnetization Process Viewpoint

When a magnetic field is applied to a material, a magnetization process takes place. This means that a change in magnetization can occur according to any of the following routes. At low applied magnetic field (and low frequency), domain boundaries bulge (this is a reversible regime, implying that if we decrease the field again, the magnetization will go back to the initial state by following the same path). This is called the domain nucleation and pinning process (see Fig. 4.4). At higher fields the walls are depinned and free to move. This is an irreversible regime (in contrast to the low-field case), meaning that, if we decrease the field, the change in magnetization upon field variation will not follow the same path. This regime is the free domain wall movement regime. At higher fields (or frequencies), the magnetization changes through the rotation of moments. In this case we have two possibilities: if the material is homogeneous and therefore behaves as if there is a single moment in the material, we get the coherent rotation regime (this is well described by the Stoner–Wohlfarth model, see Table 4.1). On the other hand, if the magnetization changes in a material in an inhomogeneous way (different points in the sample undergo

different magnetization changes), we obtain a curling process [4.26]. Finally, if the magnetic field is applied along the easy axis and then suddenly reversed, the magnetization changes with a switching process, meaning the moment jumps from a positive value to a negative value of the same magnitude without any rotation process, since the magnetization is already along the easy axis.

The Energetics Viewpoint

Hysteretic behavior can be viewed as stemming from the motion of the energy minimum as the field is varied. In Fig. 4.5 the field is cycled in a clockwise fashion. The energy minimum is displayed for several values of the field. Hysteresis is shown to arise from the asymmetric behavior of the energy with the magnetization as the field is cycled. The loss of hysteresis is given by the area circumscribed by the hysteresis loop. Since soft materials have thinner hysteresis loops (equivalent to smaller coercivity) than hard materials their losses are smaller.

The Signal Processing Viewpoint

If we apply a time-dependent field $H(t)$ to the magnetic material and monitor the magnetization $M(t)$ as the output, then the relation $M(t) = F[H(t)]$ is that of a nonlinear filter. The magnetic material acts as a nonlinear filter because M is not proportional to H except at very low fields. This nonlinearity is easily observed if $M(t)$ is monitored as a square signal with a sinusoidal input excitation $H(t)$ (Fig. 4.6). In addition, the material imposes a delay in signal propagation that is proportional to the width of the hysteresis loop (twice the coercive field). Hysteretic behavior means that different values of output are obtained when the input excitation is in-

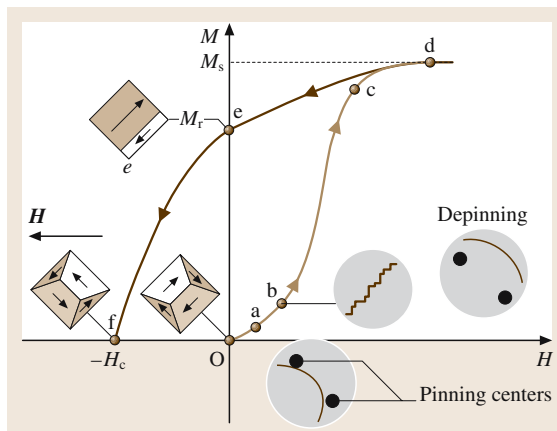


Fig. 4.4 Reversible and irreversible motion of the magnetization M with the field H . For small fields, $H < H_c$ and initial magnetization, Rayleigh's law is valid along a–b, but for large H , saturation is induced in M through irreversible domain boundary motion along b–c followed by rotation of magnetization along c–d. When the field is reversed along d–e–f towards $-H_c$, the variation in magnetization is not exactly the same as seen previously, and this is why a different $M(H)$ curve is obtained. Reversible movement at low field means that domain walls are pinned by impurities and they simply bulge under the action of H (O–a), whereas jerky irreversible movement is associated with depinning of the domain walls (a–b), creating Barkhausen noise. Free irreversible movement beyond the pinning centers is depicted on the right (after [4.29])

creased or decreased, and this phenomenon is widely exploited (in control systems for instance).

The Information Storage Viewpoint

Obtaining a nonzero response (M) at zero excitation ($H=0$) is a phenomenon known as remanence (we call this remanent value M_r). The major advantage of remanence (in addition to its usefulness in permanent magnets) is that information (the value of M_r) can be stored without any excitation source ($H=0$). Since we have M_r when the signal is decreased (we call this M_{r1}), we ought to get a different M_r when the signal is increased (we call this M_{r2}) due to hysteresis. In this case we get two bit values (0 for M_{r1} and 1 for M_{r2}). This contrasts sharply with electronic storage, where a voltage must be maintained in order to maintain the charge (representing the information); see Chapt. 51.

The Electromagnetic Compatibility and Frequency Synthesis Viewpoint

Hysteretic behavior can be viewed as the flux ϕ (corresponding to the magnetization M) induced in a circuit by an exciting current $i(t)$ at time t (corresponding to field H). The resulting relationship $\phi[i(t)]$ is a nonlinear characteristic that can be expanded as:

$$\phi[i(t)] = a_0 + a_1 i(t + \tau) + a_2 [i(t + \tau)]^2 + a_3 [i(t + \tau)]^3 + a_4 [i(t + \tau)]^4 + \dots,$$

where τ accounts for the response delay of the flux (the larger the delay, the broader the hysteresis curve). Ordinary linear inductance of the circuit corresponds to the first derivative $[d\phi/di]$, which equals a_1 for a short delay τ . Higher order terms define nonlinear inductor behavior that results in (for a short delay)

$$L(i) = a_1 + 2a_2 i + 3a_3 i^2 + 4a_4 i^3 + \dots$$

This nonlinear inductor response is interesting from a frequency synthesis point of view, since it can generate harmonics and subharmonics from the exciting current $i(t)$ (see, for instance Chua [4.30]).

Time-Dependent Viewpoint

In spite of the variety of different viewpoints describing hysteresis given above, some experimentally observed phenomena cannot be explained using any of them, including one particularly important phenomenon known as the magnetic after-effect. Experimentally, one observes a change in magnetization with time $M(t)$, despite the fact that the applied field H is kept constant, implying the presence of a time-dependent hysteresis loop stemming solely from $M(t)$. Therefore, the material must

contain units (single-domain grains) that respond differently over time, as if each unit had its own intrinsic delay; in this case the magnetization could be written in the form $M(t) = M_0 S(t) \ln(t)$, where $S(t)$ is called the magnetic viscosity. If the response time of each

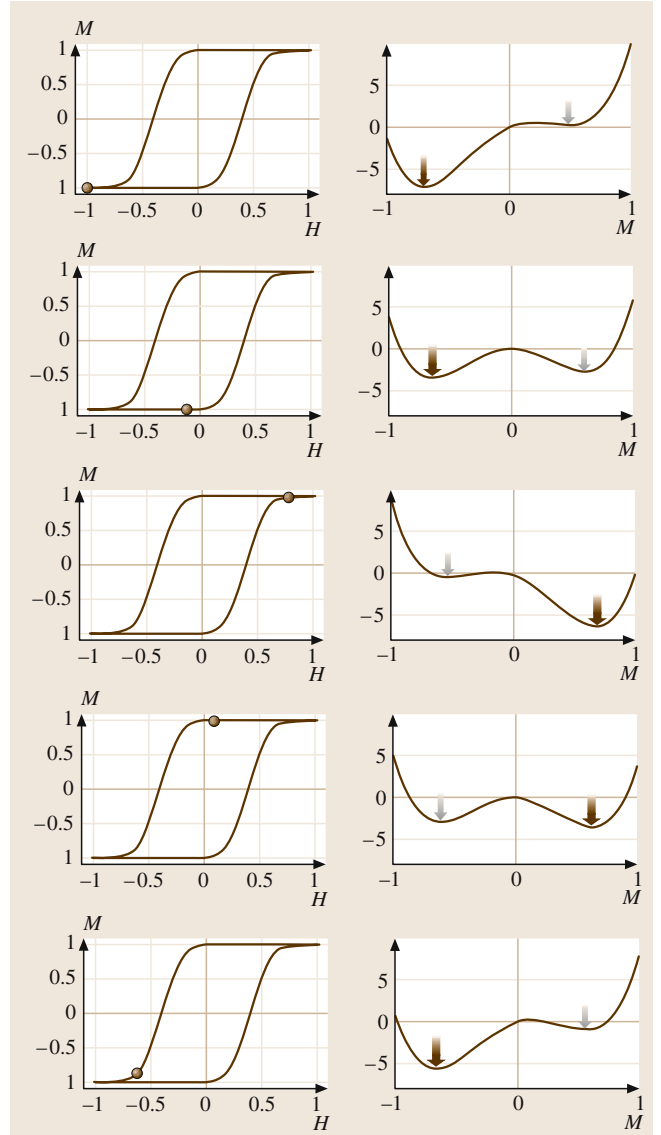


Fig. 4.5 Hysteresis loop (left) and energy as a function of magnetization (right). As the field H is varied in the hysteresis loop (as indicated by the spot ●) in a clockwise fashion, the energy changes its form and its absolute minimum (indicated by the fat arrow) moves accordingly. Hysteresis is due to asymmetric energetic behavior as the field is cycled

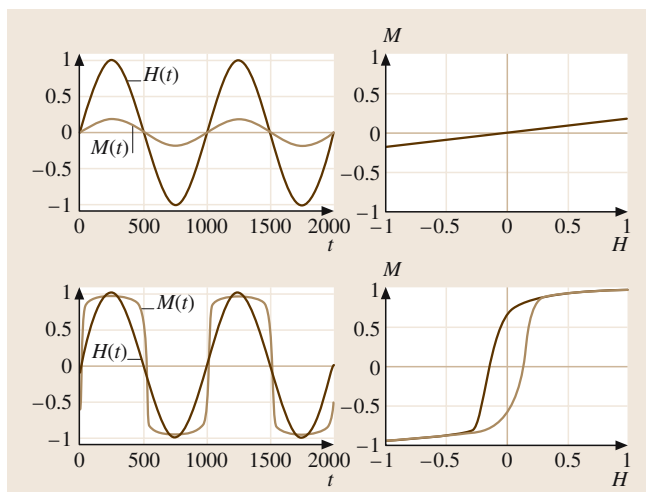


Fig. 4.6 Hysteresis loop as an input–output relationship. On the *left*, the magnetic field $H(t)$ and the magnetization $M(t)$ are displayed explicitly as functions of time t . On the *right*, the resulting input–output relationship (or M – H characteristic), the parametric curve $M(H)$, is shown. At the *top*, we have an output ($M(t)$) that is undelayed and proportional to the input $H(t)$, resulting in a paramagnetic (with no hysteresis) M – H characteristic. *Below*: an output $M(t)$ delayed and nonlinearly distorted version of the input $H(t)$, resulting in a hysteretic M – H characteristic. (After [4.31])

grain is considered to be of the thermal activation type, in other words it follows the Néel–Arrhenius formula (see Chapt. 51) $\tau = \tau_0 \exp(\Delta E/k_B T)$, where τ_0 is an “attempt time” to cross some energy barrier ΔE (see Fig. 4.5) at a temperature T , then the after-effect implies that we have a distribution of τ_0 and ΔE values. Calling the probability density function $g(\tau)$, and using a scaling initial magnetization M_0 , $M(t)$ is thus defined as:

$$M(t) = M_0 \int_0^{\infty} g(\tau) \exp(-t/\tau) d\tau.$$

In the simple case where we have a single response time τ^* , in other words $g(\tau) = 2\delta(\tau - \tau^*)$, $M(t)$ behaves as $M(t) = M_0 \exp(-t/\tau^*)$.

After analyzing the different aspects of the hysteresis loop, we now move on to investigate the physical processes that affect the magnetization behavior. We start with small magnetic fields H ; in other words $H < H_c$ (the coercive field). The initial magnetization curve follows Rayleigh’s law, $M(H) = \chi_0 H + \alpha_R H^2$, where χ_0 is the low-field susceptibility and α_R is the Rayleigh coefficient.

Since this law is valid for small fields, it describes reversible changes in the magnetization.

We now consider the physical region inside a material where the magnetization is oriented along some direction. Typically, when the extent of the material is smaller than the exchange length, one expects a single domain structure (which contains about 10^{12} – 10^{18} atoms, see also Fig. 4.1). For instance, a recording medium is considered to be made up of small grains that are made of single domains. As the recording density is increased, the grain size in the recording medium decreases, and if is small enough, its magnetization becomes sensitive to thermal energy – it can flip or reverse simply upon changes in temperature. This is called the superparamagnetic effect, and it traditionally limits longitudinal recording to 100 Gbit/in². This issue has recently been partially circumvented by introducing antiferromagnetic coupling between the storage layer and a stabilization layer. A thin film of ruthenium (7–9 Å thick) is deposited between the storage layer and the stabilization layer (Chapt. 51). If the extent of the magnetic material is large, one expects to have a multidomain structure separated by domain walls. This multidomain structure emerges in order to minimize the magnetostatic long-range dipolar interaction energy between the different moments existing in the different domains [4.32].

A domain is a piece of a material that is magnetized along a given direction (see Fig. 4.7). In general, a magnetic material contains many regions where the magnetization is aligned in a particular direction in order to minimize the magnetostatic energy. Regions with different magnetization orientations can be positioned close to one another, albeit separated by a boundary called a domain wall (typically containing about 10^2 – 10^3 atoms, Fig. 4.1). Saturation occurs when all of these regions align along some direction imposed by the external applied field. The width of a domain wall is equal (in CGS units) to $\pi(A/K)^{1/2}$, where A is the typical nearest neighbor Heisenberg exchange and K the typical anisotropy, meaning that it is the result of exchange and anisotropy; it is thinner for higher anisotropy or smaller exchange (in Fe it is about 30 nm, whereas in the hard material Nd₂Fe₁₄B it is only 5 nm). Its energy is equal to $4(AK)^{1/2}$. For bulk materials it is of the Bloch type (see Fig. 4.8), whereas for thin films (with thicknesses on the order of the exchange length) it is of the Néel type, with a width proportional to $[A/(2\pi M_s^2)]^{1/2}$. A single parameter $Q = 2K/M_s^2$ allows us to discriminate between simple ($Q < 1$) and complex wall profiles ($Q > 1$) [4.33].

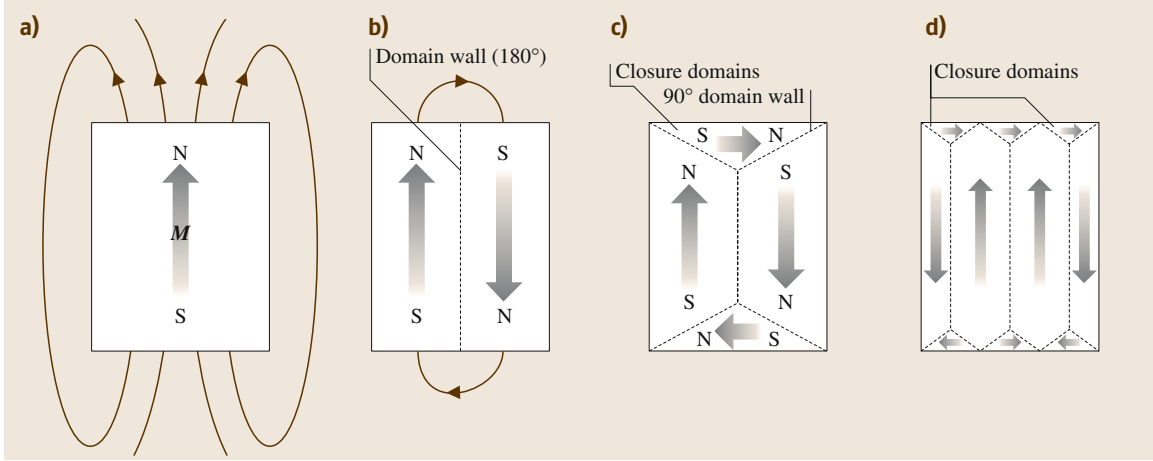
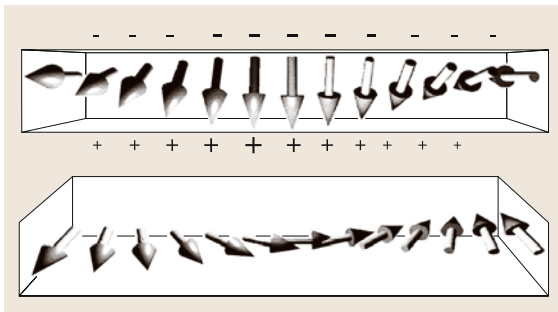


Fig. 4.7a-d Possible domain structures and wall structures in a magnetic material. **(a)** Initial magnetic configuration with a large stray field (outside the material). In order to reduce the magnetostatic energy (and the stray field), domains form, as in a highly anisotropic material **(b)**. If n domains are formed, the magnetostatic energy is reduced by n [4.27]. In a material with small anisotropy, 45° closure domains are formed to minimize the magnetostatic energy **(c)**. As more domains are formed, more walls are also formed, until a compromise is reached since it is necessary to minimize the sum of the magnetostatic energy and the wall energies **(d)**. (After [4.29])

During a magnetization process, in a hysteresis loop cycle for instance, the irreversible jerky movements of domain walls due to local instabilities created by impurities, defects, inclusions or interactions between domain walls lead to Barkhausen noise. In devices and recording media, it is better to induce magnetization changes through rotation processes because they are less noisy.

4.1.3 Intrinsic Magnetic Properties

The induction B in a linear isotropic material is related to H through the relation $B = \mu_0(H + M)$, where μ_0 is the free space permeability and M is the magnetization of the material. Using the relation $M = \chi H$, which relates the magnetization to the applied field, one gets



$B = \mu_0(1 + \chi)H$. This leads to the definition of the total permeability $\mu = \mu_0(1 + \chi)$, which yields the constitutive relation $B(H) = \mu H$. The quantity χ is called the susceptibility. The nonlinear constitutive equations $M(H)$ and $B(H)$ apply to a general nonlinear material. Such constitutive relations are similar to the $I(V)$ characteristics of electronic devices. For a linear material, μ is generally frequency-dependent. If losses occur during the propagation of an electromagnetic wave in a material due to the absorption of magnetic energy, it is possible to extend the definition of permeability to the complex plane (as in dielectrics). Losses are attributed to peaks in the imaginary part of the permeability at the absorption frequencies [4.12]. At low frequencies, losses are attributed to domains, whereas at high frequencies, losses are attributed to rotation processes. Finally, since the permeability relates two vector quantities B and H , if we consider a linear anisotropic material (such as a linear crystal), the relation $B(H) = \mu H$ becomes $B_\alpha = \mu_{\alpha\beta} H_\beta$ (Einstein summation convention used),

Fig. 4.8 Possible domain wall shapes in a magnetic material. *Above*: the Bloch type for bulk materials, with the magnetization rotating in a vertical plane with the associated poles. *Below*: the Néel type for thin films. For these walls, the magnetization rotates in a horizontal plane when the width of the film is smaller than the exchange length. (After [4.34])

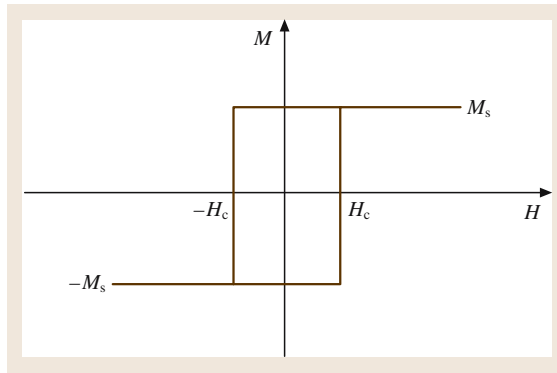


Fig. 4.9 Easy axis hysteresis loop. When the magnetic field is along the anisotropy axis (easy axis) of the material, the hysteresis loop is broadest and the magnetization jumps (switches) to $\pm M_s$ at the coercive field $\pm H_c$

and $\mu_{\alpha\beta}$ is a second-rank permeability tensor that has different components according to the relative directions of \mathbf{B} and \mathbf{H} (nine components in total, or 3^2 in three dimensions). For a nonlinear anisotropic material, we have $\mu_{\alpha\beta} = \partial B_\alpha / \partial H_\beta$.

When a material is run through a field cycle (a hysteresis loop) for the first time, the magnetization \mathbf{M} follows a path different to the $\mathbf{M}(\mathbf{H})$ lines described above. This path represents the initial magnetization, and the associated permeability is the initial permeability (Fig. 4.2). If a magnetic material is subjected to a static magnetic field \mathbf{H} and a small time-dependent field $\mathbf{h}(t)$ (alternating at frequency f) parallel to \mathbf{H} , the magnetic response is called the longitudinal permeability (at the frequency f). This kind of excitation is encountered in ferromagnetic resonance (FMR) problems (see Wigen [4.16]).

Alternatively, one can describe magnetic behavior in terms of susceptibility. This relates the magnetization to the applied field in a linear material through $\mathbf{M} = \chi \mathbf{H}$. For a nonlinear isotropic material, we have $\chi = \partial M / \partial H$, and for a nonlinear anisotropic material we have $\chi_{\alpha\beta} = \partial M_\alpha / \partial H_\beta$, with the \mathbf{M} component along the direction α and the \mathbf{H} component along β . Any jump in χ signals an important magnetic change in the material. When the temperature is decreased, the Curie temperature signals the change to ferromagnetic order from paramagnetic disorder, and the Néel temperature signals the change to antiferromagnetic order from paramagnetic disorder. Magnetic susceptibility spans several orders of magnitude (from about 10^{-5} to 10^6 cm³/mol).

The magnetization occurs along the direction that minimizes the sum of the magnetic anisotropy energy

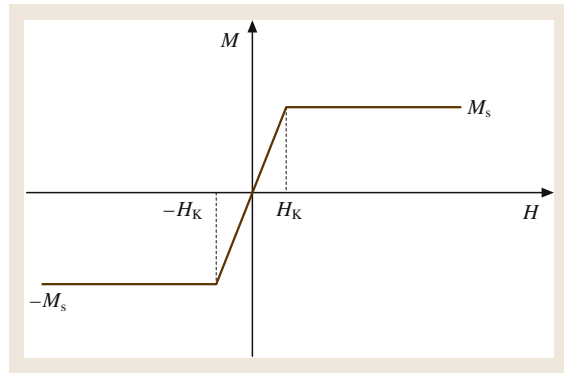


Fig. 4.10 Hard axis hysteresis loop. When the magnetic field is along the hard axis of the material, the hysteresis loop is the thinnest and the magnetization rotates smoothly from one value to another between $\pm H_K$ (the anisotropy field)

and the Zeeman energy provided by the applied field. Since the hysteresis loop changes its shape as we vary the angle of the applied field, at some point it reaches its largest width. The field then indicates the easy axis direction (also called the anisotropy axis). The extremum value of the magnetic field is the coercive field $\pm H_c$, where the magnetization (switches) jumps (see Fig. 4.9) to its saturation value $\pm M_s$.

Alternatively, let us consider the direction along which it takes most effort to align the magnetization. This is easily determined as the field direction where the hysteresis loop when it is narrowest (see Fig. 4.10). The point at which the slope changes dramatically gives the value of the anisotropy field $\pm H_K$.

Crystals are generally anisotropic because their microscopic structures are not the same in all directions. Hence, the properties (including magnetic properties) will change with the orientation of the crystal. Indeed, there will be special directions along which the magnetization prefers to point. This means that some kind of energy will be minimized when the magnetization settles along these directions: the anisotropy energy. As well as being inherent to many crystal structures, anisotropy can also be induced by an external field, a change of symmetry (for example at the surface with respect to bulk) or mechanical deformation. Microscopically, anisotropy originates from spin-orbit coupling. Coercivity increases with anisotropy; in other words the hardness increases with anisotropy (Fig. 4.11). The anisotropy energy can be evaluated from the hysteresis loop by determining the loop for various angles between the applied field and the easy axis. In general, it

is given by an expansion of the form $\sum_{\alpha,\beta} K_{\alpha\beta} M_{\alpha} M_{\beta} + \sum_{\alpha,\beta,\gamma,\delta} K_{\alpha\beta\gamma\delta} M_{\alpha} M_{\beta} M_{\gamma} M_{\delta} \dots$, where $K_{\alpha\beta}$, $K_{\alpha\beta\gamma\delta}$ are the anisotropy coefficients (second- and fourth-order respectively), and M_{α} , M_{β} are the components (direction cosines) of the normalized magnetization (by the saturation magnetization M_s). α, β, γ are the indices along the crystal x, y, z axes. This means that the norm of \mathbf{M} , $|\mathbf{M}| = (M_x^2 + M_y^2 + M_z^2)^{1/2} = 1$ in this case.

The anisotropy energy of a crystal with a single dominant axis (such as hexagonal – like cobalt, tetragonal and rhombohedral crystals) is called uniaxial. To the lowest approximation, it is given by $K_1 \sin^2 \theta$, where K_1 is the anisotropy strength, and the angle θ is the angle between the magnetization and the easy axis. The next lowest approximation allows fourth-order terms to be added to the $K_1 \sin^2 \theta$ term. In hexagonal crystals, the fourth-order term is of the form $K_2 \sin^4 \theta$. Rhombohedral symmetry allows two fourth-order terms of the form $K_2 \sin^4 \theta + K_4 \cos \theta \sin^3 \theta \cos 3\phi$. In tetragonal crystals, the fourth-order terms are of the form $K_2 \sin^4 \theta + K_4 \cos \theta \sin^4 \theta \cos 4\phi$. The easy axis is taken to be along the z -axis, the angle ϕ is the magnetization polar angle in the xy -plane, and θ is the azimuthal angle.

The anisotropy energy for a crystal with two dominant axes (including monoclinic, triclinic and orthorhombic crystals) is called biaxial. This means that two anisotropy directions exist and are competing. Some other cases are given below (see also Table 4.1 for other definitions).

Cubic Anisotropy

The anisotropy energy for the most symmetric crystals, cubic crystals, is given by $K_1 (M_1^2 M_2^2 + M_1^2 M_3^2 + M_2^2 M_3^2) + K_2 M_1^2 M_2^2 M_3^2$ (the energy is to the sixth order). Here the coefficients M_1, M_2, M_3 are the components of the normalized magnetization with respect to the crystal axes. That is, $M_1 = \sin \theta \cos \phi$, $M_2 = \sin \theta \sin \phi$, $M_3 = \cos \theta$. The easy axis, as above, is taken to be along the z -axis, and the angle ϕ is the polar angle in the xy -plane.

Helical Anisotropy

In a magnetic wire, the magnetization may prefer to lie radially in a plane perpendicular to the wire axis, resulting in a radial easy axis (radial anisotropy). If it lies along the tangents to circles that lie on planes perpendicular to the wire axis, then we have circular anisotropy. However, if it lies in a plane tangent to the lateral surface of the wire (parallel to the wire axis) and it makes

an angle that is different to 0° or 90° degrees with any plane perpendicular to the wire axis, we then have helical anisotropy.

Shape Anisotropy

A finitely sized magnetic body (an ellipsoidal shape is assumed here for simplicity) possessing uniform magnetization \mathbf{M} (represented by its components M_{α}) contains a magnetic energy (also called the magnetostatic energy) given by $2\pi N_{\alpha\beta} M_{\alpha} M_{\beta}$ (Einstein summation). The $N_{\alpha\beta}$ coefficients are the demagnetization coeffi-

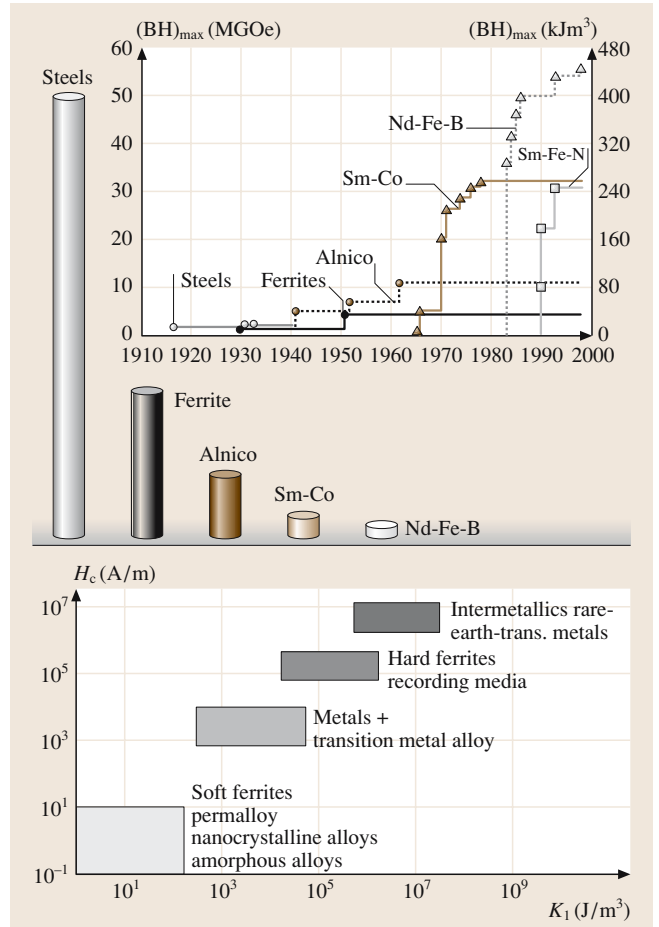


Fig. 4.11 In the upper graph, the development of hard materials for permanent magnets is associated with the value of the energy density $(BH)_{\max}$. In the middle, the progress is indicated by equivalent volume change for a fixed energy density. In the lower graph, the correlations between the coercivity H_c and the anisotropy coefficient K_1 for various materials, their magnetic hardnesses and corresponding applications. (After [4.17, 22], see Table 4.1)

cients of the body, which are determined by its shape. The origin of this terminology is its resemblance to the familiar anisotropy energy of the form $K_{\alpha\beta}\mathbf{M}_\alpha\mathbf{M}_\beta$ (Einstein summation). If a magnetic material contains N moments (atoms, ions or molecules, each carrying a moment μ_i), the energy originating from the dipolar coupling energy between the different moments is written as:

$$W_{\text{dip}} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{\mu_i \cdot \mu_j}{r_{ij}^3} - \frac{3(\mu_i \cdot \mathbf{r}_{ij})(\mu_j \cdot \mathbf{r}_{ij})}{r_{ij}^5}$$

μ_i and μ_j are two moments ($i \neq j$) in the material separated by a distance \mathbf{r}_{ij} . This energy can only be represented by constant coefficients ($N_{\alpha\beta}$) if the body has an ellipsoidal shape. Hence one can write $W_{\text{dip}} \approx 2\pi N_{\alpha\beta} \mathbf{M}_\alpha \mathbf{M}_\beta$ (Einstein convention), where the magnetization \mathbf{M} is the sum of all individual moments μ_i .

Surface Anisotropy

A finitely sized magnetic body with a bulk anisotropy (which is not too large) will realign the magnetization close to its surface in order to minimize the magnetostatic energy. In other words, the body has a surface anisotropy that is different from the bulk one. This originates from an abrupt change of symmetry at the interface between the bulk and free surface.

Anisotropy From Demagnetization

The demagnetization energy is expressed using coefficients that describe the demagnetization field \mathbf{H}_d inside a finitely sized material, created by a bulk magnetization acting against an applied external magnetic field. The components of the demagnetization field (in the ellipsoidal case) are given by (with Einstein summation) $[\mathbf{H}_d]_\alpha = -2\pi N_{\alpha\beta} \mathbf{M}_\beta$ (much like the depolarization field in the electrical case). Constant coefficients ($N_{\alpha\beta}$) are only valid when the body has an ellipsoidal shape. The coefficients depend on the geometry of the material. There are usually three positive coefficients along three directions N_{xx} , N_{yy} and N_{zz} (assuming the off-diagonal terms are all 0) in simple geometries such as wires, disks, thin films and spheres. All three coefficients are positive and smaller than 1, and their sum is 1. For a sphere, all three of the coefficients are equal to 1/3. For a thin film (or a disk) they are given by 0,0,1 when the z -axis is perpendicular to a film (or the disk) lying in the xy -plane. For a cylindrical wire of infinite length that has its axis aligned with the z -direction, the values are 1/2,1/2,0.

4.1.4 Traditional Types of Magnetism and Classes of Magnetic Materials

The main traditional types of magnetism are ferromagnetism, antiferromagnetism, ferrimagnetism, paramagnetism and diamagnetism. However, other types are also described in the review by Hurd [4.10], and with the expected advances in materials science we may expect to encounter other new classes in the future, as described in Sect. 4.2 of this work, dedicated to unconventional magnetic types (Table 4.1).

A ferromagnet is an assembly of magnetic moments interacting with a positive exchange integral that minimize their energies by adopting a common parallel configuration resulting in a net large value of total magnetization. Such a definition is valid for localized magnetism but not for itinerant ferromagnets (such as the transition metals Fe, Ni and Co), since one does not have distinct localized moments that can define an exchange integral in these materials. A ferromagnetic material (itinerant or localized) displays a characteristic hysteresis curve and remanence ($\mathbf{M} \neq 0$ for $\mathbf{H} = 0$) when one varies the applied magnetic field. When heated, the material generally loses this ordered alignment and becomes paramagnetic at the Curie temperature. Ferromagnets are usually metallic, but there are ferromagnetic insulators, such as CrBr_3 , EuO , EuS and garnets [4.27, 35].

An antiferromagnet is made up of an assembly of magnetic moments interacting with a negative exchange integral that minimize their energies by adopting an antiparallel configuration. Again, such a definition is not valid for itinerant antiferromagnets (such as Cr and Mn) since one does not have distinct localized moments and so an exchange integral cannot be defined. The net total magnetization is zero, so we do not get hysteresis. In the localized magnetism case, it is possible to consider the entire crystal as made of two interpenetrating sublattices containing moments that are all parallel inside each sublattice but where the magnetizations from all sublattices cancel each other out. When heated, the material generally loses this alternately ordered alignment and becomes paramagnetic at the Néel temperature. Oxides are generally antiferromagnetic insulators (an exception is EuO) [4.29, 35].

It is possible to have intermediate order between ferromagnets and antiferromagnets; this occurs in the ferrimagnets used in microwave devices [4.16]. In the localized case, one considers the crystal as being made of two sublattices (as in the case of an antiferromagnet), with total magnetizations that oppose one another. How-

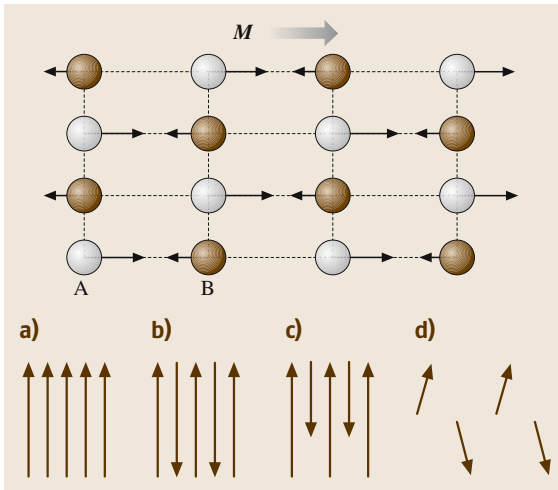


Fig. 4.12a–d Shown *above* is the magnetic moment arrangement in an ferrimagnet. The total net magnetization is nonzero. *Below*: a comparison of the magnetic arrangements between a ferromagnet (a), an antiferromagnet (b), a ferrimagnet (c) and a weak ferromagnet (d). (After [4.29])

ever, the magnitude of each magnetization is different, resulting into a net total magnetization, in contrast to an antiferromagnet.

Another intermediate case is a canted antiferromagnet, also called a weak ferromagnet. This is an assembly of magnetic moments that alternate in direction making a small inclination in the direction of the moment (canting). The small angle (Fig. 4.12d) between two neighboring moments results in a small ferromagnetic moment. Canting results from a Dzyaloshinski–Moriya exchange interaction (see Sect. 4.2.1 and [4.24, 35]). In the absence of that interaction, the moments are aligned antiparallel, like in a perfect antiferromagnet.

If we have an assembly of magnetic moments oriented at random with a zero total magnetic moment, we call it a paramagnet. Such materials do not display hysteresis or remanence when one varies the applied magnetic field. The magnetic field simply imposes some alignment onto the randomly oriented moments. For a paramagnet, the susceptibility χ is positive and many orders of magnitude smaller than for ferromagnets, and it behaves as $1/(T - T_c)$ where T_c is the Curie temperature. Some materials never order magnetically, not even at low temperatures (which implies that $T_c = 0$ K), where order prevails. On the other hand, some materials, known as superparamagnetic, show no hysteresis or remanence but have $T_c = \infty$. This does not mean that they are always ordered, but their individual moment is so large (a classical moment corresponds to a quantum angular momentum $J = \infty$, resulting in $T_c = \infty$) that it takes an infinite temperature to destroy it completely [4.27, 35].

A material that combats the influence of a magnetic field by trying to maintain zero induction ($B = 0$) while the field is applied is known as diamagnetic (copper is diamagnetic and superconductors are perfect diamagnets). The term diamagnetic is actually a misnomer, since “dia” means across, implying that the field propagates across the material, whereas the opposite is true. When a diamagnetic substance such as a silicon crystal is placed in a magnetic field, the magnetization vector \mathbf{M} in the material points in the opposite direction to the applied field. A negative susceptibility can be interpreted as the diamagnetic substance trying to expel the applied field from itself. A substance exhibits diamagnetism whenever the constituent atoms in the material have closed subshells and shells. This means that each constituent atom has no permanent magnetic moment in the absence of an applied field. Typical diamagnetic materials include covalent crystals and many ionic crystals, because the constituent atoms in these substances

Table 4.3 Examples of soft magnetic materials and their hierarchy according to the saturation magnetostriction coefficient λ_s . The main composition is successively based on Fe, NiFe and finally Co. μ_{\max} is the maximum differential permeability and H_c is the coercive field. $\text{Fe}_{80}\text{B}_{20}$ is also called Metglass 2605. $\text{Fe}_{40}\text{Ni}_{40}\text{P}_{14}\text{B}_6$ is also called Metglass 2826. The highest μ_{\max} is attained by $(\text{Fe}_{0.8}\text{Ni}_{0.2})_{78}\text{Si}_8\text{B}_{14}$, reaching 2×10^6 after annealing [4.35]

Alloy	H_c (mOe)	μ_{\max} at 50 Hz	Saturation magnetostriction coefficient
$\text{Fe}_{80}\text{B}_{20}$	40	320 000	$\lambda_s \approx 30 \times 10^{-6}$
$\text{Fe}_{81}\text{Si}_{3.5}\text{B}_{13.5}\text{C}_2$	43.7	260 000	
$\text{Fe}_{40}\text{Ni}_{40}\text{P}_{14}\text{B}_6$	7.5	400 000	$\lambda_s \approx 10 \times 10^{-6}$
$\text{Fe}_{40}\text{Ni}_{38}\text{Mo}_4\text{B}_{18}$	12.5–50	200 000	
$\text{Fe}_{39}\text{Ni}_{39}\text{Mo}_4\text{Si}_6\text{B}_{12}$	12.5–50	200 000	
$\text{Co}_{58}\text{Ni}_{10}\text{Fe}_5(\text{Si}, \text{B})_{27}$	10–12.5	200 000	$\lambda_s \approx 0.1 \times 10^{-6}$
$\text{Co}_{66}\text{Fe}_4(\text{Mo}, \text{Si}, \text{B})_{30}$	2.5–5	300 000	

have no unfilled sub-shells. Since the diamagnetic material tries to minimize the effect of \mathbf{H} , it expels field lines – a phenomenon that can be exploited in magnetic levitation. A superconductor is a perfect diamagnet, and a metal exposed to high frequencies is partially diamagnetic, since the applied field can only penetrate it to skin depth. The susceptibility χ is constant for a diamagnet (it does not vary with temperature) and is slightly negative. Superconductors have $\chi = -1$ (below critical temperature), whereas semiconductors have the following values of susceptibility (in cm^3/mol) at room temperature: Si, -0.26×10^{-6} ; Ge, -0.58×10^{-6} ; GaAs, -1.22×10^{-6} ; as given by Harrison [4.36].

Materials with a relatively small coercive field (typically smaller than 1000 A/m), as preferred in transformer cores and magnetic read heads, are called soft magnetic materials. Most (but not all) simple metals, transition metals and their compounds are soft. Permalloys, amorphous and nanocrystalline alloys and some ferrites are soft. Amorphous materials are soft because their disordered structure does not favor any direction (no anisotropy energy), whereas nanocrystals possess anisotropy over a short length scale (although it can be larger than its bulk counterpart). Softness is also measured by the maximum permeability attainable (see Table 4.3 of soft elements).

On the other hand, a material with a relatively large coercive field (typically larger than 10 000 A/m), as preferred in permanent magnets, motors and magnetic recording media (disks and tapes), is called a hard magnetic material. This means that stored data is not easily lost since a large field is required to alter the magnitude of magnetization. Most (but not all) rare-earth metals, their compounds and intermetallics are hard. There are also hard ferrites. Permanent magnets are used in power systems (in power relays), motors and audio/video equipment (such as headphones, videotapes); see Table 4.1 and Fig. 4.11.

Ferrites are ferrimagnetic ceramic-like alloys. During the early development of ferrites, the compositions of all ferrites could be described as FeOFe_2O_3 . However, more modern ferrites are better described as MOFe_2O_3 , where M is a divalent metal (note that the Fe in Fe_2O_3 is trivalent). Ferrites are used in microwave engineering and recording media due to their very low eddy current losses [4.12] and the fact that they operate over a large frequency interval (kHz to GHz). The ratio of the resistivity of a ferrite to that of a typical metal can reach as high as 10^{14} . They are made by sintering a mixture of metallic oxides MOFe_2O_3 where $\text{M}=\text{Mn}, \text{Mg}, \text{Fe}, \text{Zn}, \text{Ni}, \text{Cd}...$ Ferrite read heads are however limited to

frequencies below 10 MHz as far as switching is concerned, and this is why several new types of read head (thin films, AMR, GMR, spin valves, magnetic tunnel junctions) have been or are being developed in order to cope with faster switching (Chapt. 51). Other conventional magnetic materials similar to ferrites include spinels and garnets, as described below.

Spinel

Spinel are alloys with the composition $(\text{MO})_x(\text{MO})_{1-x}\text{Fe}_2\text{O}_3$ (a generalization of the ferrite composition), that have the structure of MgAl_2O_4 (which provides the origin of the word spinel) [4.16].

Garnets

Garnets are oxides that have compositions related to the spinel family, of the form $(3 \text{M}_2\text{O}_3, 5 \text{Fe}_2\text{O}_3)$, that crystallize into the garnet cubic structure $[\text{Ca}_3\text{Fe}_2(\text{SiO}_4)_3]$. They are ferromagnetic insulators of general formula $\text{M}_3\text{Fe}_5\text{O}_{12}$, where M is a metallic trivalent ion ($\text{M}=\text{Fe}^{3+}$ for example). Garnets have been used in memory bubble technology, lasers and microwave devices (because their ferromagnetic resonance linewidth with respect to the field is small, on the order of a fraction of an Oersted, when the resonance frequency is several tens of GHz [4.16].), especially those of general formula $\text{M}_3\text{Ga}_5\text{O}_{12}$. For instance, when produced as a thin film a few microns thick, $\text{Gd}_3\text{Ga}_5\text{O}_{12}$, called GGG (gadolinium gallium garnet), exhibits perpendicular anisotropy with domains (bubbles) that have up or down magnetization (perpendicular with respect to the film plane). Thus, a bit can be stored in a bubble and can be controlled using a small magnetic field. GGG is considered to be one of the most perfect artificially made crystals, since it can be produced with extremely few defects (less than 1 defect per cm^2). Another nomenclature, called the [cad] notation, is used with rare-earth iron garnets of general formula $\text{X}_3\text{Y}_2\text{Z}_3\text{O}_{12}$. This notation means: dodecahedral (c site is surrounded by 12 neighbours and represented by element X); octahedral (a site is surrounded by 8 neighbours and represented by element Y); and tetrahedral (d site is surrounded by 4 neighbours and represented by element Z). The most important characteristic of these garnets is the ability to adjust their compositions and therefore their magnetic properties according to selected substitutions on the c, a or d sites. The element X is a rare earth, whereas Y and Z are Fe^{3+} . The magnetization is changed by placing nonmagnetic ions on the tetrahedral d site: increasing the amount of Ga^{3+} , Al^{3+} , Ge^{4+} , or Si^{4+} will decrease the magnetization. On the other hand, increas-

ing the amount of Sc^{3+} or In^{3+} at the octahedral a site will increase the magnetization. Ion substitution can also be used to tailor other magnetic properties (including anisotropy, coercivity and magnetostriction). Garnets are typically grown using liquid phase epitaxy at a growth speed that easily reaches a micron in thickness in one minute, and a very high yield is achieved. These materials have not only been used in

bubble materials but also in magneto-optical displays, printers, optical storage, microwave filters and integrated optics components. Despite all of these attractive properties, their easy tunabilities and very high yields, problems soon arose with the limited access times of bubble memories, since the switching frequency was found to be limited to less than about 10 MHz [4.16] (Chapt. 51).

4.2 Unconventional Magnetism

4.2.1 Conventional and Unconventional Types of Exchange and Coupling in Magnetic Materials

In conventional magnetic materials, the strength of the magnetic interactions between two neighboring localized moments i and j (as in atoms/ions/molecules and rare-earth solids) is described by a direct exchange interaction. The latter is essentially a Coulomb (electrostatic) interaction between the electrons at the i and j sites. The word exchange is used because the corresponding overlap integral describing this interaction involves wavefunctions with permuted (exchanged) electron coordinates (in order to respect the Pauli exclusion principle). The exchange energy between sites i and j is given by $-A_{ij}\mathbf{M}_i\cdot\mathbf{M}_j$ where \mathbf{M}_i and \mathbf{M}_j are, respectively, the magnetization at the i and j sites [4.29]. When $A_{ij} > 0$, the energy is minimized when the moments are parallel, and when $A_{ij} < 0$ an antiparallel configuration of the moments is favored. The energy arising from exchange over a distance r in the continuum limit is approximately AM^2/r^2 .

The RKKY (Ruderman–Kittel–Kasuya–Yoshida, [4.27]) oscillatory interaction occurs between two localized moments mediated by a surrounding electron gas. It varies in 3-D systems as $\cos(2k_F r)/r^3$, where r is the distance between the moments and k_F is the Fermi wavevector of the electron gas. It was recently discovered that a counterpart of the RKKY interaction exists in 2-D between two magnetic thin films separated by a metallic spacer [4.37]. The RKKY-like interaction between the two magnetic films across a metallic spacer is oscillatory with respect to the spacer thickness z . Thus it becomes possible to decide to couple the magnetic films positively (ferro) or negatively by changing the thickness z of the metallic spacer (see Fig. 4.1 for typical lengths). This is extremely useful for thin film devices (see Chapt. 51 and Table 4.1).

The Dzyaloshinski–Moriya exchange interaction is a vectorial exchange interaction between two neighboring localized moments (\mathbf{M}_i and \mathbf{M}_j) of the form $D_{ij}\mathbf{M}_i\times\mathbf{M}_j$, which contrasts with the scalar ordinary exchange interaction of the form $A_{ij}\mathbf{M}_i\mathbf{M}_j$. This cants (producing a small inclination between) two neighboring antiferromagnetic moments that are usually antiparallel, resulting in weak ferromagnetism. This is due to asymmetric spin-orbit effects [4.24].

Present interest is focused on magnetic thin films and their interactions. Information storage, sensing, spintronics, quantum computing and other applications of magnetic thin film devices are the main spur to understand the nature and extent of magnetic exchange interactions and coupling effects as well as those that arise between magnetic 2-D layers in order to tailor appropriate devices. Novelty is expected since the device size is comparable to the interaction length (Fig. 4.1).

The coupling strength of the interaction between two neighboring magnetic films i and j can be modeled by a factor J_{ij} . This is similar to the exchange integral A_{ij} between two neighboring localized moments, but it involves entire layers generally made from itinerant magnets and not the single moments used with A_{ij} . The exchange interaction is of the form $J_{ij}\mathbf{M}_i\cdot\mathbf{M}_j$, where \mathbf{M}_i and \mathbf{M}_j are the magnetizations per unit surface of films i and j . The main interest in J_{ij} lies in the fact that its range is longer in reduced dimensions (1-D and 2-D) than in 3-D (for instance, an RKKY-like interaction between two magnetic films across a metallic spacer is oscillatory with a longer range than it is in 3-D, since it varies like $1/r^2$ instead of $1/r^3$), and its physics is entirely different from the standard RKKY interaction between localized moments [4.37]. The sign of the interaction depends on the thickness of the metallic spacer. Other types of exchange exist between films, such as biquadratic or higher order with a generalized Heisenberg form $I_{ij}[\mathbf{M}_i\mathbf{M}_j]^n$ where $n \geq 2$ and I_{ij} is a coupling constant

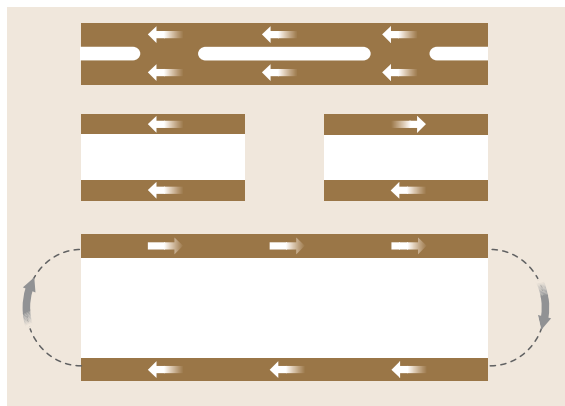


Fig. 4.13 Coupling and exchange between magnetic layers separated by a nonmagnetic (metallic or insulating) spacer. The *top* drawing shows ferromagnetic coupling arising from pinholes that exist across the spacer, whereas the *middle* drawing shows exchange-coupled layers that might be ferro- (*left*) or antiferromagnetic (*right*), depending on the (generally metallic) spacer thickness. Finally, the *bottom* drawing schematizes magnetostatic dipolar coupling between the top and bottom layers. (After [4.10].)

that is a function of n and the separation between the films (Table 4.1).

Interactions between two ferromagnetic (or other types of magnetic) films are not limited to metallic spacers; they can occur across an insulating or semiconducting spacer (Fig. 4.13). Magnetic pinholes generally mediate direct ferromagnetic coupling between neighboring multilayers. They are viewed as shorts across the insulating or semiconducting spacer. Néel proposed that conformational roughness (the orange-peel model, see *Schulthess and Butler* [4.38] for instance) at interfaces can result in ferromagnetic coupling for a moderate thickness of spacer material. Magnetostatic dipolar coupling occurs between the roughness features or between domain walls in the two magnetic layers. In this case, stray flux fields from walls in one film can influence the magnetization reversal process in the other. The Slonczewski loose-spin model (see Table 4.1, related to exchange and coupling in multilayers) is based on the change in angular momentum experienced by spin-polarized electrons tunneling across an insulating spacer, resulting in a magnetic exchange coupling. The magnitude and sign of the coupling oscillates depending on the interfacial barrier properties ([4.10] and Table 4.1).

In the absence of a spacer, the exchange energy per unit surface area of ferromagnetic and antiferromagnetic films gives rise to a new phenomenon called exchange

bias (the shift of the hysteresis curve of the ferromagnetic film, see entry in Table 4.1). An important consequence of this bias is a pinning-down of the direction of the ferromagnetic moment, which is used in low-noise read heads because it hinders noisy (jerky) domain wall motion (Chapt. 51). Exchange bias is also called anisotropic exchange, and is still poorly understood from a fundamental level despite the fact it was discovered more than 50 years ago.

4.2.2 Engineering and Growth of Thin Magnetic Films

A magnetic film is considered to be thin if its thickness is smaller than the (bulk-defined) exchange length; in other words its thickness corresponds to the spatial extent of a single domain [4.10]. Magnetic thin films possess very attractive and distinct physical properties with respect to their bulk counterparts. For instance, reduced dimensionality, coordination and symmetry lead to magnetocrystalline anisotropy energies that are two or three orders of magnitude larger than for the bulk (the phenomenon of quenching of the orbital angular momentum – as seen in transition metal ions – is absent, see *Kittel* [4.27]). Saturation magnetization M_s is also enhanced with respect to bulk, since the effective Bohr magneton number n_B per atom/molecule is larger at lower dimensions (for instance in nickel clusters n_B can reach 1.8, whereas in the bulk it is only 0.6). Couplings are also enhanced in thin films, as in the case of the RKKY-like interaction across a metallic spacer. Coupling between two magnetic layers across a semiconducting spacer or an insulator leads to entirely new interactions such as those described by the Slonczewski loose-spin model or the Bruno quantum interference model. New couplings also arise, such as anisotropic exchange or exchange bias between a ferro and an antiferromagnetic layer, and new types of multilayers can be made (spring magnets) by alternating hard and soft thin films (Table 4.1).

When we consider the growth of magnetic materials as thin films, we should recall that these films possess distinct physical properties with respect to the materials used in conventional microelectronics, such as surface energies. The free energies γ of the surface and the interface play dominant roles. They determine the growth modes in thermal equilibrium; the morphology of material B grown on material A depends on the balance between the free surface energies of the substrate, the overlayer and the interface [4.10]. Transition metal-based magnetic materials exhibit a relatively

Table 4.4 Surface energies γ (in J/m²) for magnetic and nonmagnetic materials, listed with respect to their atomic number for the low-energy cleavage surface. These are approximate values, which are difficult to measure in general and depend on surface orientation and reconstruction [4.10]

Magnetic metal γ (J/m ²)	Cr 2.1	Mn 1.4	Fe 2.9	Co 2.7	Ni 2.5	Gd 0.9				
Transition metal γ (J/m ²)	Ti 2.6	V 2.9	Nb 3.0	Mo 2.9	Ru 3.4	Rh 2.8	Pd 2.0	Ta 3.0	W 3.5	Pt 2.7
Simple or noble metal γ (J/m ²)	Al 1.1		Cu 1.9		Ag 1.3		Au 1.6			
Semiconductor γ (J/m ²)	Diamond 1.7		Si 1.2		Ge 1.1		GaP 1.9		GaAs 0.9	
Insulator γ (J/m ²)	LiF 0.34		NaCl 0.3		CaF ₂ 0.45		MgO 1.2		Al ₂ O ₃ 1.4	

high surface energy, owing to their partially filled d shells. Noble metal substrates have smaller surface energies, and insulating substrates even less. Additionally, when one performs epitaxial growth, another concern is lattice-matching the different underlayers, as displayed in Table 4.4 and Table 4.5, along with typical quantities of interest in representative magnetic materials.

4.2.3 Electronic Properties: Localized, Free and Itinerant Magnetism and Spin-Polarised Band Structure

Building a working device requires an understanding of not only magnetic properties but also electronic properties and their interplay. We expect that new devices will be constructed from a variety of magnetic (conventional and unconventional) materials as well as others already known in microelectronics. Insulating oxides (except EuO) and rare-earth compounds with well-localized external-shell f electrons are solid state

materials with atom-like magnetism. Magnetic atoms, ions and molecules or associated with well-defined localized orbitals and individual moments arising from orbital, spin or total angular momentum. When these moments get close to one another, as in the solid state, they interact as defined by Heisenberg: $A_{ij}\mathbf{M}_i \cdot \mathbf{M}_j$ [4.29]. The latter is altered by the presence of the surrounding free-electron gas. Therefore we must understand magnetism in a free-electron gas, its counterpart arising from localized moments, and finally its nature when we have the hybrid case (itinerant magnetism), which occurs in a transition metal (free s and localized d electrons). This problem is very complicated, and so we will rely upon a “one-electron approximation” of band structure, and more precisely spin-polarized band structure [4.11].

In a free-electron gas, one can assume independent noninteracting electrons, so many-electron and nonlocal effects (arising from exchange) do not need to be taken into account. Magnetism in this case is due to individual electron spins and it is straightforward to establish that so-called Pauli paramagnetism holds [4.27]. In addition,

Table 4.5 Lattice-matched combinations of magnetic materials, substrates and spacer layers. There are two main groups of lattice-matched systems with lattice constants close to 4.0 Å or 3.6 Å respectively, after making 45° rotations of the lattice or doubling the lattice constant (After *Himpsel et al.*[4.10] with minor editing)

First group					
Magnetic metal $2^{1/2}a$ (Å) (a (Å))	Cr (bcc) 4.07 (2.88)	Fe (bcc) 4.05 (2.87)	Co (bcc) 3.99 (2.82)		
Simple or noble metal a (Å)	Al 4.05	Ag 4.09	Au 4.07		
Semiconductor $a/2^{1/2}$ (Å) (a (Å))	Ge 3.99 (5.65)	GaAs 4.00 (5.65)	ZnSe 4.01 (5.67)		
Insulator a (Å) (a/2 ^{1/2} (Å))	LiF 4.02 (2.84)	NaCl 5.65 (3.99)	MgO 4.20 (2.97)		
Second group					
Material a (Å)	Fe (fcc) 3.59	Co (fcc) 3.55	Ni (fcc) 3.52	Cu 3.61	Diamond 3.57

orbital effects give rise to Landau diamagnetism. The spin-polarized density of states is shifted by the field contribution $\pm\mu_B B$ depending on the spin orientation with respect to the field B . In the free-electron Pauli case, the magnetization can be evaluated (at $T = 0$ K) from

$$\begin{aligned} M &= \mu_B (N_{\uparrow} - N_{\downarrow}) \\ &= \frac{\mu_B}{2} \int_{-\mu_B B}^{E_F} dE N(E + \mu_B B) \\ &\quad - \frac{\mu_B}{2} \int_{\mu_B B}^{E_F} dE N(E - \mu_B B) \end{aligned}$$

which yields $M \approx \mu_B^2 N(E_F) B$, where μ_B is the Bohr magneton, $N(E)$ is the free electron and $N_{\uparrow\downarrow}(E)$ the spin-polarized density of states. The contribution from Landau (orbital) diamagnetism is a lot more complicated to evaluate, but it is exactly minus one third of the paramagnetic expression [4.27].

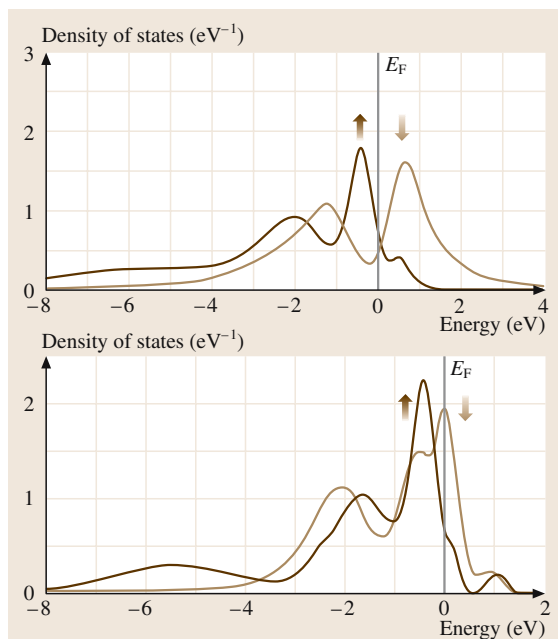


Fig. 4.14 Spin-polarized density of states for bcc Fe and fcc Ni. The bands for the spin up electrons are separated from the spin down band by the exchange splitting energy. The method used is superior to the **LSDA** method which yields a noisy density of states. The authors call it **LDA+DMFT** (local density approximation + dynamical mean field theory). (After [4.39])

States derived from the 3d and 4s atomic levels are responsible for the physical properties of transition metals (itinerant magnets). The 4s electrons are more spatially extended (have a higher principal quantum number) and determine (for instance) the compressibility, whereas the 3d states determine the magnetic properties. The 3d electrons still propagate throughout the material; hence the term itinerant magnetism. The spin-polarized band structure of a transition metal is different to that of a conventional electronic material (see Fig. 4.14 and Fig. 4.15 for the electronic band structures of iron, nickel and cobalt and Table 4.6 for their physical properties).

Many metals have an odd number of electrons (in the atomic state), and so one would expect them to be magnetic due to their unpaired spins; however, only five of them are actually magnetic (Cr, Mn, Fe, Ni and Co, which are all transition metals with s–d hybridization). Note that copper is diamagnetic, with $\chi = -0.77 \times 10^{-6} \text{ cm}^3/\text{mol}$ at room temperature; Fe, Ni and Co are ferromagnets; whereas Cr and Mn are antiferromagnets.

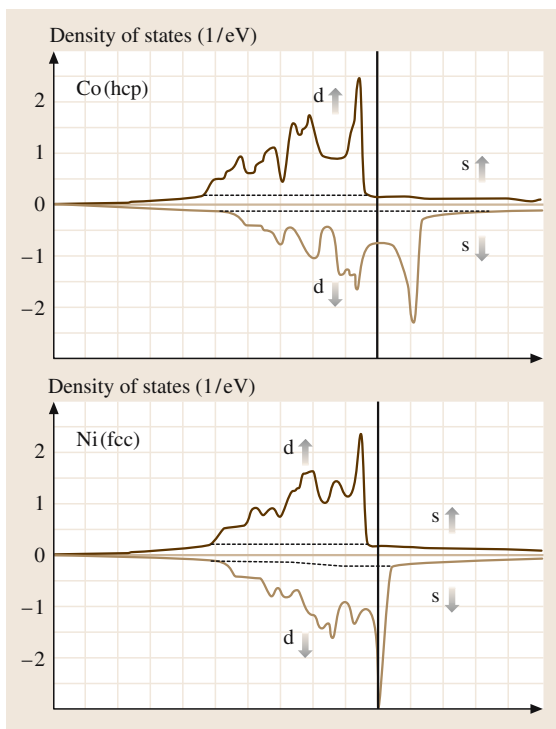


Fig. 4.15 Symmetric double-sided representation of the spin-polarized density of states for Co and Ni (obtained with the LMTO method or linearized muffin-tin orbitals), with the Fermi level indicated. (After [4.40])

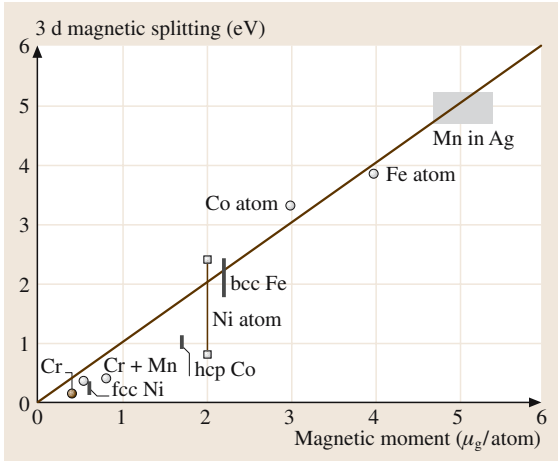


Fig. 4.16 Exchange splitting versus the magnetic moment (in Bohr magnetons μ_B per atom) in 3d transition metals. The diagonal line is 1 eV/ μ_B , and the magnetic moment values are given for the element (in the atomic state) and the corresponding crystal. (After [4.10])

It is not straightforward to model magnetism and related properties in transition metal alloys due to the itinerant nature of the electrons. In solids, magnetism arises mainly from electrostatic (Coulomb) electron–electron interactions, namely the exchange interactions, and in magnetic insulators these can be described rather simply by associating electrons with particular atomic

Table 4.6 Fundamental magnetic properties of the transition metals Fe, Co and Ni [4.35]. Note that the domain wall thicknesses and energies are approximate

Magnetic property	Fe	Co	Ni
n_B at $T = 0$ K (in Bohr magnetons)	2.22	1.72	0.62
M_s at 300 K (in 10^6 A/m)	1.71	1.40	0.48
Exchange energy A (meV)	0.015	0.03	0.020
Curie temperature (K)	1043	1404	631
Anisotropy K_1 (J/m^3) at 300 K	4.8×10^4	45×10^4	-0.5×10^4
Lattice spacing (nm)			
a	0.29	0.25	0.35
c		0.41	
Domain wall thickness			
• in nm	40	15	100
• in lattice parameters	138	36	285
Domain wall energy (J/m^2)	3×10^{-3}	8×10^{-3}	1×10^{-3}

sites so that Heisenberg exchange $A_{ij}\mathbf{M}_i \cdot \mathbf{M}_j$ can be used to describe the behavior of these systems. In metallic systems it is not possible to distribute the itinerant electrons in this way, and such simple pairwise interactions between sites cannot be defined. Metallic magnetism is a complicated many-electron effect that has required significant effort over a long period to understand and describe it [4.10]. A widespread approach is to map this problem onto one involving independent electrons moving in the fields set up by all of the other electrons. It is this aspect that gives rise to the spin-polarized band structure that is often used to explain the properties of metallic magnets (such as the non-integer values of multiples of the Bohr magneton μ_B). In addition, compositional structure plays a major role in itinerant magnetism. Consider for instance the ordered Ni–Pt alloy. This is an antiferromagnet, whereas its disordered counterpart is ferromagnetic. It can be said that itinerant magnetism occurs between two extreme limits: the localized and the completely free electron cases (more appropriate to alkali metals), as described below. As far as transport is concerned, one can use the Stoner two-band model, which is a two-fluid model (as used for electrons and holes in semiconductors) with one population of electrons with spin up and another with spin down, along with the corresponding spin-polarized densities of states (Fig. 4.14, Fig. 4.15 and Himpfel et al. [4.10].)

In ordinary crystalline nonmagnetic materials, the electronic band structure is built from the lattice (periodic) potential seen by a single electron (with any spin state) in the structure. Mathematically, this amounts to solving the Schrödinger equation corresponding to that potential and calculating its eigenfunctions (spin-independent wavefunctions). The eigenvalues are the ordinary bands. In a magnetic material, one must include exchange effects (which derive from Pauli exclusion principle-dependent electrostatic interactions that yield non-local interactions), and one must account explicitly for spin in the wavefunctions. Several methods exist to do this, such as the LSDA (local spin density approximation, see Fig. 4.1) which is a spin extension of the LDA (local density approximation) where one builds a local approximation to the exchange interaction from the local density of electrons [4.39]. The local spin density approximation (see Fig. 4.1) formalism provides a reliable description of the magnetic properties of transition metal systems at low temperatures. It also provides a mechanism for generating non-integer effective numbers n_B of Bohr magnetons, together with a plausible account of the many-electron nature of magnetic moment forma-

tion at $T = 0$ K. The bands obtained depend on spin, as depicted in the figures cited earlier. One can obtain the spin-dependent band structure from the spin-polarized density of states for each spin polarization (up \uparrow or down \downarrow , also called the majority and minority, like in ordinary semiconductors). Tables 4.7 and 4.8 give the polarizations and some spin-dependent band-structure quantities for representative transition metals and their alloys.

This semiconductor-like nomenclature (majority–minority) will eventually become confusing when we start dealing with metals and semiconductors simultaneously. For the time being, however, this nomenclature is acceptable so long as we are dealing solely with magnetic metals, and one can define a new type of gap (originating from the exchange interaction) called the exchange splitting or (spin) gap between two spin-dependent bands (Fig. 4.14 and Fig. 4.15). This explains the existence of novel materials such as half-metals, which can be contrasted with semi-metals (graphite) where we have a negative electronic gap because of valence and conduction band overlap.

Half-metals (such as CrO_2 and NiMnSb) possess one full spin-polarized band (up for instance) while the other (down) is empty. These materials are very important for spintronics and (especially) when injecting spin-polarized carriers.

4.2.4 Prospects for Spintronics and Quantum Information Devices

Presently, we are witnessing the extension of electronics to deal with spin and charge instead of charge only, the realm of traditional electronics. The reason spin becomes interesting and useful stems from the following ideas. As device integration increases and feature length decreases towards the nanometer scale, the spins of individual carriers (electrons or holes) become good quantum numbers. This means that spin value is conserved over the nanoscale (the spin diffu-

sion length is typically 5–50 nm), whereas it was not previously (in the micron regime), so it can be used in the nanometer regime to carry useful information. This means that carriers transport energy, momentum, charge and additionally spin. In addition, there is the potential that quantum computers could be constructed from spintronic components [4.25]. In perfect analogy with ordinary electronics, spintronics is based on four pillars:

1. Spin injection: How do we create a non-equilibrium density of spin-polarized carriers – electrons with spin up n_\uparrow (or down n_\downarrow) or holes with spin up p_\uparrow (or down p_\downarrow)? This can be viewed as the spin extension of the Haynes–Shockley experiment, and it can be done optically or electrically using thin magnetic layers serving as spin filters or analyzers/polarizers as in spin valves (Table 4.1). Spin injection can also be achieved with carbon nanotubes since they do not alter the spin state over large distances (Chapt. 51).
2. Spin detection: How do we detect the spin and charge of a non-equilibrium density of spin-polarized carriers?
3. Spin manipulation: How do we alter and control the spin and charge states of a non-equilibrium density of spin-polarized carriers?
4. Spin coherence: How do we maintain the spin and charge states of a non-equilibrium density of spin-polarized carriers over a given propagation length?

Spintronics is based, like microelectronics, on particular materials, their growth techniques (epitaxial or other) and their physical properties (electrical, mechanical, magnetic, thermal), as well as the ability to fabricate thin films and objects of reduced dimensionality (quantum dots, quantum wires, quantum pillars, clusters) and a knowledge of different processing steps (oxidation, diffusion, doping, implantation, etching, passivation, thermal insulation, annealing, texturing, sputtering, patterning) that can be used to build useful devices. All of the techniques established in the

Table 4.7 Spin polarization expressed in % for several ferromagnetic materials according to several authors. The discrepancies between the different results stem from the various approaches used to estimate the density of states at the Fermi level, and points to how difficult it is to obtain a unanimous figure. Note that for a half-metal like CrO_2 or NiMnSb , the polarization is 100%

	<i>Meservey, Tedrow [4.41]</i>	<i>Moodera, Mathon [4.42]</i>	<i>Monsma, Parkin [4.43]</i>
Fe	40	44	45
Co	35	45	42
Ni	23	33	31
$\text{Ni}_{80}\text{Fe}_{20}$	32	48	45
$\text{Co}_{50}\text{Fe}_{50}$	47	51	50

Table 4.8 Magnetic splittings δk_{ex} , full width half maxima δk_{\uparrow} and δk_{\downarrow} , and spin-dependent mean free paths λ_{\uparrow} and λ_{\downarrow} for NiFe and NiCr alloys ($\pm 0.01 \text{ \AA}^{-1}$) (After [4.44])

	δk_{ex} (\AA^{-1})	δk_{\uparrow} (\AA^{-1})	δk_{\downarrow} (\AA^{-1})	λ_{\uparrow} (\AA)	λ_{\downarrow} (\AA)
Ni	0.14	0.046	0.046	> 22	> 22
Ni _{0.9} Fe _{0.1}	0.14	0.04	0.10	> 25	10
Ni _{0.8} Fe _{0.2}	0.14	0.03	0.22	> 33	5
Ni _{0.93} Cr _{0.07}	0.09	0.096	0.086	11	10
Ni _{0.88} Cr _{0.12}	≤ 0.05	0.12	0.11	8	9

field of microelectronics must be enlarged and extended to magnetism-based electronics (magnetoelectronics), which highlights the challenge of controlling magnetic interactions that have an anisotropic vectorial character in contrast to the scalar electrical interactions (based solely on charge) present in conventional microelectronic devices. Magnetoelectronics introduces the notions of anisotropy (magnitude, nature and direction), coercivity, saturation magnetization, and so on, which need to be controlled during growth (magnetic field-assisted growth or epitaxial growth must further developed in order to favor magnetic anisotropy along desired directions). For instance, there is, in some devices, the need to grow amorphous metallic magnetic layers (in order to get a very small anisotropy, resulting in a very magnetically soft layer), and these may be harder to grow than amorphous semiconductors or insulators (in metals, a cooling speed of one million degrees per second is typically needed to get an amorphous material). Temperature is a very important

parameter, since ferromagnetism (antiferromagnetism) is lost above the Curie (Néel) temperature. New types of materials emerge when spin-polarized carriers are used: for instance, half-metallic materials (such as CrO₂ and NiMnSb) that possess carriers that are completely polarized in terms of spin (all up or all down). Additionally, magnetic interactions can be based on localized, free, itinerant, para-, dia-, ferro-, antiferro- or ferrimagnetic materials, which can be electrically metallic, insulating or semiconducting. For instance, the possibility of controlling ferromagnetic interactions between localized spins in a material using transport carriers (electrons or holes), as well as the demonstration of efficient spin injection into a normal semiconductor, have both recently renewed the interest in diluted magnetic semiconductors. If made functional at a reasonably high temperature, ferromagnetic semiconductors would allow one to incorporate spintronics into usual electronics, which would pave the way to integrated quantum computers (Table 4.1).

References

- 4.1 R. Boll, H. Warlimont: IEEE Trans. Magn. **17**, 3053 (1981)
- 4.2 E. C. Stoner, E. P. Wohlfarth: Phil. Trans. R. Soc. London A **240**, 599 (1948)
- 4.3 C. L. Platt, M. R. McCartney, F. T. Parker, A. E. Berkowitz: Phys. Rev. B **61**, 9633 (2000)
- 4.4 J. C. Slonczewski: Phys. Rev. B **39**, 6995 (1989)
- 4.5 N. C. Koon: Phys. Rev. Lett. **78**, 4865 (1997)
- 4.6 C. Tannous, J. Gieraltowski: J. Mater. Sci. Mater. El. **15**, 125 (2004)
- 4.7 R. White: IEEE Trans. Magn. **28**, 2482 (1992)
- 4.8 F. Schatz, M. Hirscher, M. Schnell, G. Flik, H. Kronmüller: J. Appl. Phys. **76**, 5380 (1994)
- 4.9 M. J. Dapino, R. C. Smith, F. T. Calkins, A. B. Flatau: J. Intel. Mat. Syst. Str. **13**, 737 (2002)
- 4.10 F. J. Himpsel, J. E. Ortega, G. J. Mankey, R. F. Willis: Adv. Phys. **47**, 511 (1998)
- 4.11 H. F. Jansen: Physics Today (Special Issue on Magnetoelectronics) **April**, 50 (1995)
- 4.12 J. B. Goodenough: IEEE Trans. Magn. **38**, 3398 (2002)
- 4.13 H. J. Richter: J. Phys. D **32**, 147 (1999)
- 4.14 P. Farber, M. Hörmann, M. Bischoff, H. Kronmüller: J. Appl. Phys. **85**, 7828 (1999)
- 4.15 P. Coeure: J. Phys. (Paris) Coll. C-6 **46**, 61 (1985)
- 4.16 P. E. Wigen: Thin Solid Films **114**, 135 (1984)
- 4.17 O. Gutfleisch: J. Phys. D **33**, 157 (2000)
- 4.18 G. Burkard, D. Loss: Europhys. News **Sept-Oct**, 166 (2002)
- 4.19 H. Hauser, L. Kraus, P. Ripka: IEEE Instru. Meas. Mag. **June**, 28 (2001)
- 4.20 G. A. Prinz: Science **282**, 1660 (1998)
- 4.21 I. Zutic, J. Fabian, S. Das Sarma: Rev. Mod. Phys. **76**, 323 (2004)

- 4.22 H. Kronmueller: J. Magn. Magn. Mater. **140–144**, 25 (1995)
- 4.23 J. L. Simonds: Physics Today (Special Issue on Magnetoelectronics) **April**, 26 (1995)
- 4.24 C. M. Hurd: Contemp. Phys. **23**, 469 (1982)
- 4.25 M. A. Nielsen, I. L. Chuang: *Quantum Computation and Quantum Information* (Cambridge Univ. Press, New York 2000)
- 4.26 S. Chikazumi: *Physics of Ferromagnetism*, Int. Ser. Monogr. Phys., 2nd edn. (Oxford Univ. Press, Clarendon 1997)
- 4.27 C. Kittel: *Introduction to Solid State Physics*, 6th edn. (Wiley, New York 1986)
- 4.28 M. R. Fitzsimmons, S. D. Bader, J. A. Borchers, G. P. Felcher, J. K. Furdyna, A. Hoffmann, J. B. Kortright, I. K. Schuller, T. C. Schulthess, S. K. Sinha, M. F. Toney, D. Weller, S. Wolf: J. Magn. Magn. Mater. **271**, 103 (2004)
- 4.29 S. O. Kasap: *Principles of Electronic Materials and Devices*, 3rd edn. (McGraw-Hill, New York 2001)
- 4.30 L. O. Chua: *Introduction to Non-Linear Network Theory* (McGraw-Hill, New York 1969)
- 4.31 B. K. Chakrabarti, M. Acharyya: Rev. Mod. Phys. **71**, 847 (1999)
- 4.32 A. Hubert, R. Schäfer: *Magnetic Domains* (Springer, Berlin, Heidelberg 1998)
- 4.33 A. P. Malozemoff, J. C. Slonczewski: *Magnetic Domains in Bubble-Like Materials* (Academic, New York 1979)
- 4.34 D. Buntinx: . Ph.D. Thesis (Université Catholique de Louvain, Louvain 2003)
- 4.35 D. Jiles: *Introduction to Magnetism and Magnetic Materials*, 2nd edn. (Chapman and Hall, New York 1991)
- 4.36 W. A. Harrison: *Electronic Structure and the Properties of Solids* (Freeman, New York 1980)
- 4.37 P. Bruno: Phys. Rev. B **52**, 411 (1995)
- 4.38 T. C. Schulthess, W. H. Butler: J. Appl. Phys. **87**, 5759 (2000)
- 4.39 A. I. Lichtenstein, M. I. Katsnelson, G. Kotliar: Phys. Rev. Lett. **87**, 067205 (2001)
- 4.40 A. Barthelemy: GDR Pommes Proceedings **CNRS** publication (June, Aspet, France, 2001)
- 4.41 R. Meservey, P. M. Tedrow: Phys. Rep. **238**, 173 (1994)
- 4.42 J. Moodera, G. Mathon: J. Magn. Magn. Mater. **200**, 248 (1999)
- 4.43 D. J. Monsma, S. S. P. Parkin: Appl. Phys. Lett. **77**, 720 (2000)
- 4.44 K. N. Altmann, N. Gilman, J. Hayoz, R. F. Willis, F. J. Himpsel: Phys. Rev. Lett. **87**, 137201 (2001)

5. Defects in Monocrystalline Silicon

The aggregation of intrinsic point defects (vacancies and Si interstitials) in monocrystalline silicon has a major impact on the functioning of electronic devices. While agglomeration of vacancies results in the formation of tiny holes (so-called “voids”, around 100 nm in size, which have almost no stress field), the aggregation of Si interstitials exerts considerable stress on the Si matrix, which, beyond a critical size, generates a network of dislocation loops around the original defect. These dislocation loops are typically microns in size. Consequently, they are much more harmful to device functioning than vacancy clusters. However, the feature size in electronic devices has now shrunk down to the 100 nm scale, meaning that vacancy aggregates are also no longer acceptable to many device manufacturers.

This chapter is intended to give an introduction to the properties of intrinsic point defects in silicon and the nucleation and growth of their aggregates. Knowledge in this field has grown immensely over the last decade. It is now possible

5.1	Technological Impact of Intrinsic Point Defects Aggregates	102
5.2	Thermophysical Properties of Intrinsic Point Defects	103
5.3	Aggregates of Intrinsic Point Defects	104
5.3.1	Experimental Observations	104
5.3.2	Theoretical Model: Incorporation of Intrinsic Point Defects	107
5.3.3	Theoretical Model: Aggregation of Intrinsic Point Defects	109
5.3.4	Effect of Impurities on Intrinsic Point Defect Aggregation	112
5.4	Formation of OSF Ring	115
	References	117

to accurately simulate the aggregation process so that the defect behavior of semiconductor silicon can be precisely tailored to the needs of the device manufacturer. Additionally, the impact of various impurities on the aggregation process is elucidated.

As the feature size continues to shrink in device industry, a thorough understanding of defect behavior in bulk silicon becomes more and more important. Three major defect types relevant to device performance have been identified: vacancy aggregates (known as “voids”, which usually have a size of less than 150 nm); Si interstitial clusters embedded in a network of dislocation loops, each of which extend over several microns (L-pits); and large grown-in oxygen precipitates. The latter generate stacking faults (OSF) during wafer oxidation. The voids form in the center of the crystal, while L-pits are observed in the outer region. The two concentric defect regions are usually separated by a small OSF ring. The type of defect that develops in the growing crystal is determined by a simple parameter: the ratio of

the pull rate to the temperature gradient at the growth interface. In industry, crystals with only one type of defect – voids – are produced almost exclusively. The formation and behavior of voids has been studied intensively and is accurately described by current theoretical models. As the feature size is now approaching the void size, the growth of so-called “perfect silicon” with almost no detectable defects may be adopted. Furthermore, the doping of crystals with impurities like nitrogen or carbon is being widely investigated. These impurities can significantly reduce the defect size, but they may also have harmful effects, such as enhancing the generation of OSFs. Some models have recently been proposed which may allow us to predict some of the effects of impurities.

5.1 Technological Impact of Intrinsic Point Defects Aggregates

Single intrinsic point defects in silicon – vacancies and interstitials – have not been found to have any negative impact on device performance so far. However, if they aggregate into clusters they can be even detrimental to device functionality. This is also true of extended defects like dislocations. When silicon wafer technology was in its formative years, much of the work devoted to improve wafer quality focused on controlling dislocation density in the silicon crystals, as it was not possible to grow dislocation-free crystals. However, with the introduction of dislocation-free crystal growth processes into mass production, the issue of extended dislocations in relation to bulk silicon quality vanished. As the feature size decreased and the demand for higher device performance increased, it soon became apparent that intrinsic point defects and their aggregation during the cool-down phase of the crystal growth process were having an increasingly negative impact on device performance and yield. Historically, one of the first serious challenges in this regard was the aggregation of Si interstitials in floating zone (FZ)-grown crystals, which results in a local network of dislocation loops (secondary defects), so-called “A-swirls” [5.1] or L-pits [5.2]. Although the diameters of these dislocation loops are only a few microns, they are large enough to generate hot spots in the space charge regions of high-power devices [5.3, 4]. In the second half of the 1980s, the industry began to encounter problems with the early breakdown of the gate oxide in memory devices based on Czochralski (CZ)-grown silicon [5.5]. After intensive gate oxide integrity (GOI) investigations, it was found that the root cause of the gate oxide degradation was tiny micro holes – voids – which were formed by vacancy aggregation during crystal growth [5.6, 7]. Each void is thermally stabilized by an oxide layer present on its inner surface. After wafer polishing, the voids show up as dimples or laser light scattering (LLS) defects on the wafer surface, causing a local thinning of the gate oxide [5.8, 9]. Voids are considerably smaller (less than 150 nm) than A-swirl defects, and so their impact on device performance is only apparent if the location of a void coincides with that of an active element, such as a transistor. In addition, most of these defective transistors can be repaired due to the built-in redundancy of memory chips. Consequently, vacancy aggregates are tolerable for many devices, so long as their density is insignificant compared to those of device process-induced defects. This should be contrasted with A-swirls or L-pits, which al-

ways result in permanent device damage due to their large sizes.

Empirically, it has been found that gate oxides 40–50 nm thick are most susceptible to void defects [5.9]. Thinner oxides show higher GOI yields, and when the thickness drops below 5 nm the influence of voids on the GOI yield disappears [5.10, 11]. However, as the feature size continues to shrink, additional adverse effects have been identified, such as shorts between trench capacitors and lack of device reliability [5.12, 13]. As the design rule becomes equal to or less than the void size, these problems are expected to aggravate and device manufacturers may have to switch to materials with extremely small defect sizes or those that contain virtually no defects.

There are three main ways to achieve silicon with no harmful intrinsic point defect aggregates. The first is to grow silicon crystals in a regime where Si interstitials and vacancies are incorporated in equal concentrations (see Sect. 5.3.2), resulting in almost complete mutual annihilation of point defects (so-called “perfect” or “pure” silicon) [5.14–17]. The method inevitably involves lower pull rates and very tight control over crystal growth parameters, which yields considerably lower throughput and higher costs, in particular for 300 mm crystal growth.

The second approach is the growth of nitrogen-doped crystals with very fast pull rates (high cooling rates) and subsequent high-temperature ($\approx 1200^\circ\text{C}$) wafer annealing [5.18]. Nitrogen doping in conjunction with a fast pull rate decreases the vacancy aggregate size (Sect. 5.3.3 and Sect. 5.3.4) [5.19, 20], meaning that they are easy to dissolve using a high-temperature wafer treatment. Void annihilation first requires the dissolution of the inner oxide layer, which in turn, necessitates the outdiffusion of oxygen. Thus, annealed wafers only exhibit a near-surface defect-free region $\approx 10\mu\text{m}$ in depth, which is, however, sufficient for device manufacturing. Annealed wafers also take advantage of the notable mechanical strengthening effects of nitrogen doping [5.21–23], which helps to suppress slippage generation during high-temperature treatment. One very recent development is rapid thermal wafer annealing at a temperature of around 1300°C . At this temperature, outdiffusion of oxygen is not necessary because the oxygen concentration is usually below that required for oxygen solubility, so the inner oxide layer dissolves throughout the bulk and the voids collapse. This process yields silicon of a simi-

lar quality to that resulting from perfect silicon crystal growth.

The third method is the well-known epi wafer approach.

All of these methods require rather precise defect engineering in order to obtain the properties demanded by the device industry, except in the case of pp+ epi wafers.

Here, the high boron concentration of the substrate suppresses intrinsic point defect aggregation (Sect. 5.3.4) and enhances oxygen precipitation in the bulk. Therefore, this wafer type not only provides a defect-free epi layer, but also gives metallic contaminants superior internal gettering (impurity removal) abilities and high slip resistances.

5.2 Thermophysical Properties of Intrinsic Point Defects

Understanding intrinsic point defect aggregation undoubtedly requires rather exact knowledge of their respective thermophysical properties.

The intrinsic point defects – vacancies and Si interstitials – can exist in different configurations. Generally, six localized point defect configurations of high symmetry are considered: the vacancy and the split-vacancy on the one hand, and the tetrahedral, the hexagonal, the bond-centered and the [100] split or dumbbell Si interstitial on the other [5.24]. While the localized configuration works rather well for vacancies, theoretical calculations strongly favor an extended configuration of lower symmetry for Si interstitials [5.25, 26]. The extended self-interstitial model was originally proposed to explain the high pre-exponential factor in the coefficient of self-diffusion, and this model now has support from theoretical calculations [5.24, 27, 28]. According to theory, the high value of the pre-exponential factor results from the multitude of self-interstitial configurations with similar energies and the significant lattice relaxations that accompany some of these configurations.

Vacancies and Si interstitials can also exist in various charged states (such as V^{2+} , V^+ , V^0 , V^- , V^{2-}), and at the high temperatures ($> 1000^\circ\text{C}$) where point defects start to aggregate all states should be present in dynamic equilibrium [5.29]. Due to this equilibration, it is not meaningful to assign a specific charge to vacancies and Si-interstitials, respectively. However, atomistic calculations show that the charged states are much higher in energy and so their populations should be negligible. So far, there is no indication that charged states have any impact on defect aggregation and so they are not considered in current defect nucleation models.

Unfortunately, it is generally not possible to observe intrinsic point defects directly and so their thermophysical properties cannot be measured directly either. Thus, indirect approaches must be used that involve fitting defect concentrations along with many other parameters. Various experimental systems have been used to

infer the thermophysical properties of point defects. The most common of these are metal diffusion experiments where a metallic contaminant such as zinc, gold or platinum is introduced into the bulk via high-temperature drive-in diffusion [5.30–32]. The diffusion rate of the metallic impurity, which is easily detectable using standard methods, is related to the mobility and the concentration of intrinsic point defects (kick-out and Frank–Turnbull mechanism), which provides a way to indirectly probe the behavior of the point defects. These experiments yield good estimates for the product $C_{\text{eq}}D$ of the equilibrium concentration C_{eq} and the diffusivity D for self-interstitials I and vacancies V, respectively. The following values are derived from zinc diffusion results [5.33]:

$$C_{\text{I}}^{\text{eq}} D_{\text{I}} = 1.5 \times 10^{26} \exp(-4.95 \text{ eV}/k_{\text{B}}T) \text{ cm}^{-1} \text{ s}^{-1}$$

and

$$C_{\text{V}}^{\text{eq}} D_{\text{V}} = 1.3 \times 10^{23} \exp(-4.24 \text{ eV}/k_{\text{B}}T) \text{ cm}^{-1} \text{ s}^{-1}.$$

Another frequently used experimental method is the defect analysis of CZ crystals grown with varying pull rates. In this case, the observables are the dynamic responses of the oxidation-induced stacking fault (OSF) ring and the interstitial–vacancy boundary as a function of changes in crystal growth process conditions [5.2, 34–36] (Sect. 5.4). These observables have been quantitatively correlated to intrinsic point defect distributions in crystals and, can therefore be used to derive thermophysical properties [5.37]. Of particular importance is the complementary nature of crystal growth and metal diffusion experiments with regard to parametric sensitivity. The high temperature dependence of the IV–boundary and its sensitivity to self-interstitial and vacancy competition implies that these experiments are suitable for determining some pre-exponential coefficients. The metal diffusion experiments, which can be carried out over a wide range of temperatures, are particularly useful for probing activation energies.

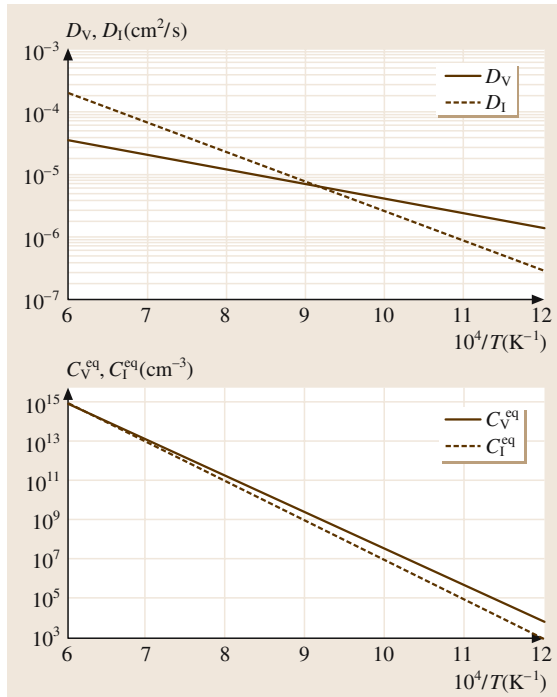


Fig. 5.1 Diffusivities and equilibrium concentrations of vacancies and Si interstitials, respectively. (After [5.41])

In recent years, RTA experiments with fast cooling rates have been carried out with silicon wafers to investigate the influence of the free vacancy concentration on oxygen precipitation [5.17, 38]. By heating the wafers to above 1150 °C, detectable concentrations of Frenkel pairs are created in the wafer bulk. The wafer is then held at this temperature for around 10 s to allow the self-interstitials to diffuse out, resulting in a near-equilibration of both interstitials [$C_I^{\text{eq}}(T)$] and vacancies [$C_V^{\text{eq}}(T)$]. Following this rapid equilibration, the wafer is rapidly cooled (at rates of ≈ 100 °C/s) to freeze in the vacancy excess which can subsequently be probed via its impact on oxygen precipitation. This technique is especially useful for gaining information on [$C_V^{\text{eq}}(T) - C_I^{\text{eq}}(T)$] at various temperatures [5.39]. Quenching experiments at different holding tempera-

tures are also reported for hydrogen-doped samples, which allow us to quantitatively measure the VH_4 complex [5.40]. From these results, the vacancy formation energy was deduced as ≈ 4 eV with relatively good accuracy.

Experimental data that relate to the recombination coefficient of vacancies and Si interstitials are scarce [5.42, 43]. Nevertheless, there is general agreement that recombination between vacancies and Si interstitials is so fast that the product of the concentrations of both species is always in thermal equilibrium. The exact value is therefore not relevant to subsequent discussion here.

There have been numerous attempts to compute the thermophysical properties of point defects directly by atomistic simulations. The most important aspect of these atomistic calculations is the accuracy of the assumed interatomic interaction. As the ab initio approach which explicitly considers electronic interactions is extremely computationally expensive and therefore limited to small systems and zero temperature, most simulations are carried out using empirical potentials, such as the Stillinger–Weber or Tersoff potential functions, discarding explicit representations of electronic interactions. An excellent overview of the current status of this field is given by Sinno [5.44]. A parameter set that yields rather good results for defect aggregation is shown in Fig. 5.1. It is clear that the difference between the vacancy and Si interstitial concentrations at the melting point is only around 30% in favor of vacancies. On the other hand, vacancies diffuse more slowly than Si interstitials. These two properties are of great relevance to the peculiar defect behavior of silicon.

In the past, fitted values for equilibrium concentrations and diffusivities inferred from various experimental data often spread over many orders of magnitude, meaning that they were of little help. However, in recent years, theoretical methods and the accuracy of experimental data have significantly improved and the predicted values are now in much better agreement with those gleaned from experiments. Today, the uncertainties in thermophysical data are usually less than 10%.

5.3 Aggregates of Intrinsic Point Defects

5.3.1 Experimental Observations

As pointed out in Sect. 5.1, the dominant intrinsic point defect aggregates in silicon single crystals are voids

and L-pits/A-swirls. Figure 5.2 shows a TEM image of a void with a 5 nm oxide layer on the inner surface. Oxidation of the inner surface is not observed in FZ crystals due to their very low oxygen contents. The

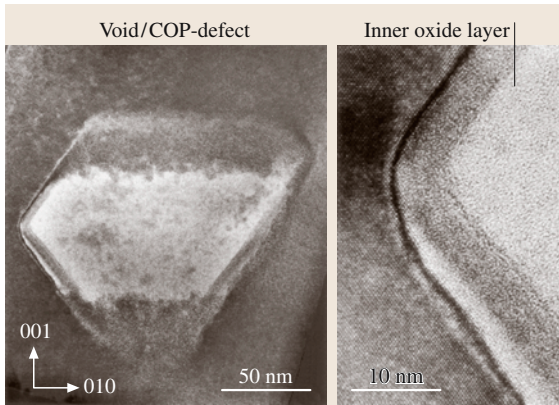


Fig. 5.2 TEM picture of a void/COP (left). The inner surface of the void is covered by an oxide layer (right). (After [5.52])

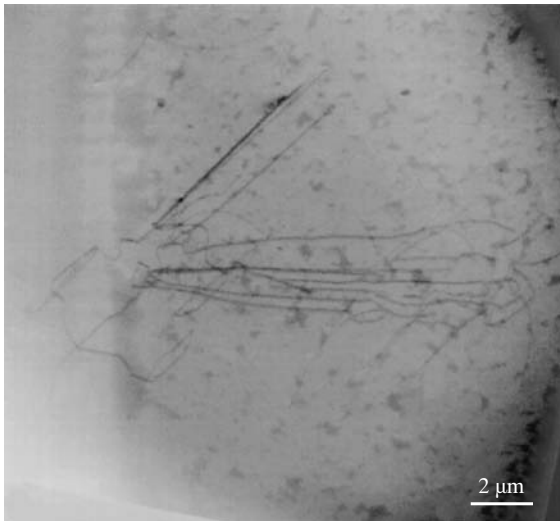


Fig. 5.3 TEM picture of an L-pit. (After [5.53])

thickness of the oxide layer decreases with increasing cooling rate of the growing crystal. A strain field is generally not observed around the void. Usually, voids exhibit octahedral morphologies, but they can change to platelet or even rodlike shapes if the crystal is doped with nitrogen (Sect. 5.3.4). Under standard growth conditions, twin voids consisting of two partial octahedral voids are predominantly observed [5.45, 46]. At lower oxygen contents and higher cooling rates of the growing crystals, single octahedral voids are preferentially formed [5.47, 48]. Typical sizes range from 70–200 nm depending on the crystal growth conditions. Historically, different notations were introduced for this type of defect

depending on their delineation or the detection technique used: crystal originated particle (COP) [5.49], laser light scattering tomography defect (LSTD) [5.2], flow pattern defect (FPD) [5.50], and D-defects [5.51]. Today, it is generally accepted that they all denote the same defect type: vacancy aggregates.

The TEM image of an L-pit is depicted in Fig. 5.3. The network of perfect dislocation loops characteristic of this defect type is clearly visible. The core of the defect is a self-interstitial aggregate which forms an extrinsic stacking fault [5.53, 54]. When the stacking fault reaches a critical size, the strain exerted on the crystal lattice is relaxed by the generation of dislocation loops. It is believed that the B-swirl observed in FZ

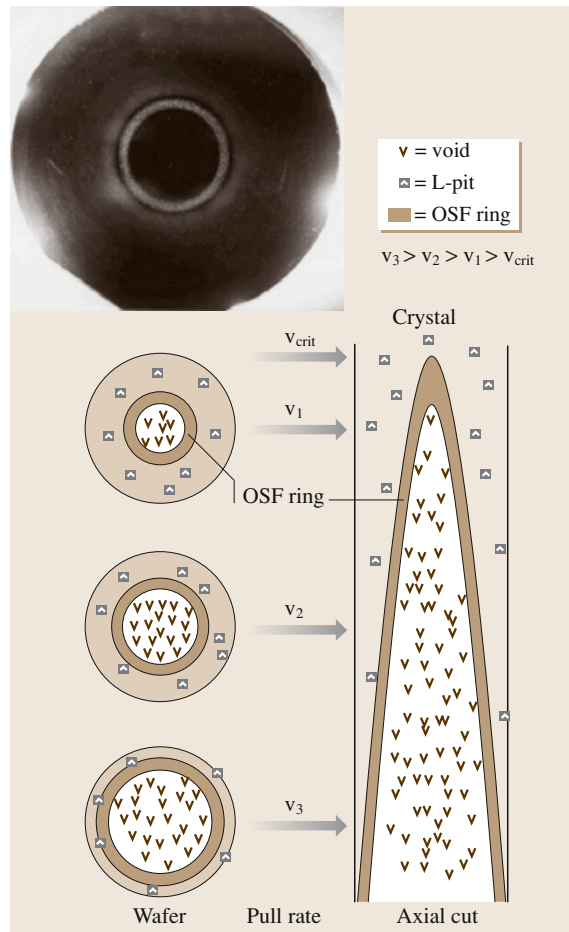


Fig. 5.4 Variation in the radial defect behavior of a CZ crystal as a function of pull rate. The insert on the left shows a wafer from the crystal part grown with medium pull rate. The defects were delineated by Secco etching

crystals [5.55,56] is most likely related to self-interstitial aggregates with sizes below the critical limit [5.57,58]. While the B-swirl can be annihilated by appropriate high-temperature treatment, the L-pit/A-swirl is extremely stable.

In the case of epi wafers, voids at the substrate surface are covered by the epilayer and no defects are generated in the epilayer. However, if the layer thickness is very thin ($< 1 \mu\text{m}$), a dimple remains at the epilayer surface which is detectable by laser light scattering [5.60]. In contrast to voids, L-pits always generate clearly visible and harmful defects in the epilayer [5.61] and are therefore unacceptable to device manufacturers. It is well-established that the pull rate V and the axial temperature gradient G at the growth interface of the growing crystal have the biggest influence on the defect types that develop in the growing crystal [5.34,35,62] and their spatial [5.36] as well as their density/size distributions [5.59]. At high pull rates, vacancy-related defects (voids) are observed over the entire crystal volume (Fig. 5.4). When the pull rate is reduced, oxidation-induced stacking faults (OSFs) develop in a small ring-like region near the crystal rim [5.63]. If the pull rate is decreased further, the ring diameter shrinks and L-pits are detected in the area outside the OSF ring [5.2]. No voids are found in this outer region. Thus, the OSF ring obviously represents the spatial boundary between vacancy- and self-interstitial-type defects. At a critical pull rate, the void region and the OSF ring disappear completely and only L-pits are detected. In FZ crystals with their inherently low oxygen contents, no OSF ring is observed as a boundary; instead a defect free-zone is observed as the boundary [5.64].

The radial position of the OSF ring can be approximately described by the equation [5.36]:

$$V/G(r) = \xi_{\text{tr}} = 1.34 \times 10^{-3} \text{ cm}^2 \text{K}^{-1} \text{min}^{-1},$$

Here r is the radial position. Thus, the parameter V/G controls the type of grown-in defect: if $V/G > \xi_{\text{tr}}$ vacancy aggregates develop, while for $V/G < \xi_{\text{tr}}$ Si interstitial-related defects are observed. This was first recognized by Voronkov in 1982 [5.62].

A further important parameter for the control of grown-in defects is the thermal history of the crystal. While its influence on the type of defect that develops in the growing crystal is negligible, detailed investigations have revealed a strong effect of cooling rate on the defect density/size distribution. For a crystal cooling in the temperature range between roughly 1000°C and 1100°C , there seems to be an exponential relation between the void density N_{void} and the dwell time t

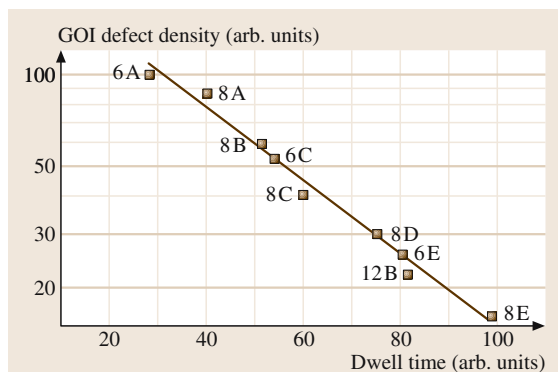


Fig. 5.5 Void density as derived from GOI measurements as a function of the dwell time of the growing crystal at $\approx 1100^\circ\text{C}$. The straight line denotes simulation results. The various data points relate to different growth processes. (After [5.59])

(Fig. 5.5) [5.65]. The defect density data in Fig. 5.5 were derived from GOI measurements that can detect much smaller defect sizes than delineation techniques based on etching and/or laser light scattering. The correlation between the various detection methods down to LSE sizes of $0.12 \mu\text{m}$ was found to be rather good [5.19].

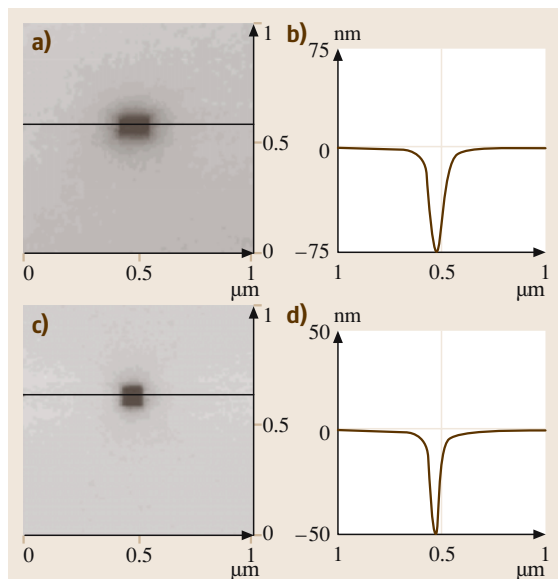


Fig. 5.6 Atomic force microscopy measurements of voids at the wafer surface. The depth of the surface void is rather close to its original size in the crystal. Samples are from a slowly cooled (left) and a rapidly cooled crystal (right). (After [5.60])

The exact upper value of the above temperature range depends on the crystal growth conditions and will be shown (Sect. 5.3.3) to coincide with the void nucleation temperature. The width of the selected temperature interval, on the other hand, does not have much impact on the above empirical relation. As will be discussed below, the relevant temperature interval is rather small [5.66]. Hence, the dwell time might just as well be replaced by the inverse of the cooling rate at the nucleation temperature.

It was further observed that the average void size increases as dwell time increases (with slower cool-

ing rates) and vice versa (Fig. 5.6) [5.59, 67]. Another striking feature is the characteristic change in the defect density/size distribution across the crystal diameter (Fig. 5.7) [5.68]. While large voids of low density prevail in the crystal center, the distribution gradually shifts to small sizes and high density towards the boundary of the void region. As the cooling rates of the crystal center and rim are almost the same, this remarkable variation in the density/size distribution must be related to the radial inhomogeneity of $V/G(r)$. Unfortunately, similar data are not yet available for L-pits. The behavior of this defect type is more complex, as small interstitial aggregates which do not generate dislocation loops are not detected as L-pits.

Valuable information about the defect nucleation temperature is obtained from transient growth experiments where the pull rate is reduced to zero and the growth process is halted when a certain crystal length has been reached [5.69–71]. After a certain amount of time has elapsed, growth is resumed and the crystal is grown to full length. During the halting period, the various crystal positions are at different temperatures according to the axial temperature gradient of the crystal. By analyzing the axial defect behavior of the as-grown crystal, one can identify the temperature ranges where specific defect types nucleate and grow in size. The results of those experiments clearly demonstrated that the nucleation temperature of voids lies between 1000°C and 1100°C and varies with the growth conditions. It was also demonstrated that the nucleation and growth period of the voids occurs over a temperature interval of less than 50 K [5.72]. Void growth is then obviously stopped by the formation of the inner oxide layer. The oxide layer continues to grow down to relatively low temperatures. The scarce data available for L-pits indicate a similar temperature range for nucleation [5.73].

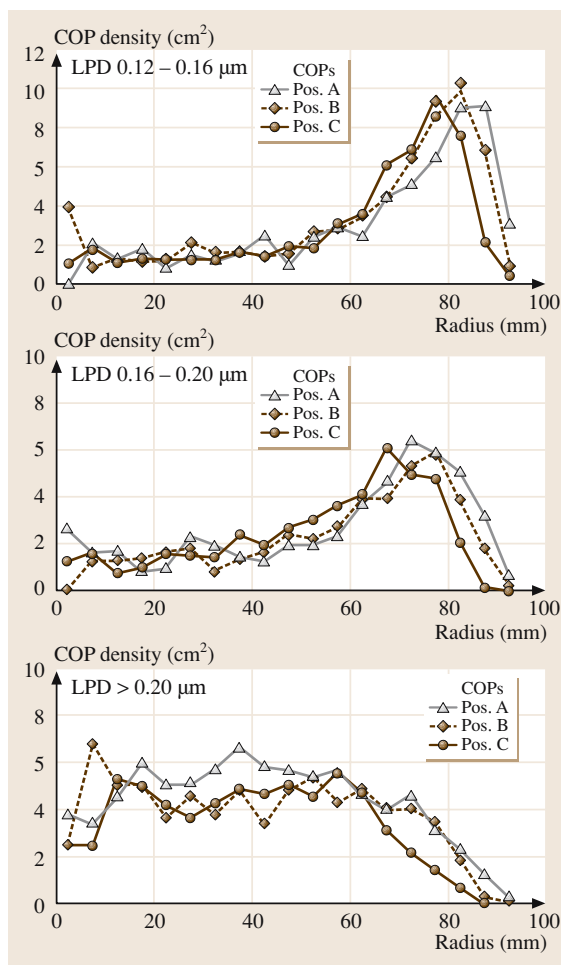


Fig. 5.7 Radial variation in the density/size distribution of voids/COPs across a wafer as measured by laser light scattering. The light point defect size (LPD) correlates well to the actual void size. The positions A, B and C relate to different axial crystal positions

5.3.2 Theoretical Model: Incorporation of Intrinsic Point Defects

The incorporation of intrinsic point defects is the result of two competing fluxes at the growth interface: a vacancy and an interstitial flux which are both directed into the crystal (Fig. 5.8). Each of the two fluxes consists of two components. The first flux component is driven by the advancing growth interface, which generates a convective flux proportional to the pull rate V , and the second is driven by the vacancy/interstitial recombination behind the interface, which creates a concentration gradient and, in turn, a Fickian diffusion flux. The latter is proportional to G . If the pull rate is low,

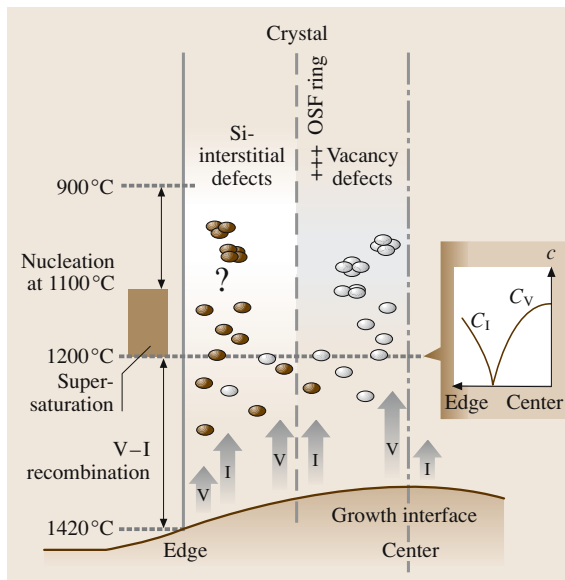


Fig. 5.8 Schematic picture of the incorporation of vacancies (V) and Si interstitials (I), respectively, into the growing crystal. The insert on the right shows the radial variation of the remaining species after V-I recombination has ceased and supersaturation starts

then the Fickian diffusion dominates over the convective flux. As self-interstitials diffuse significantly faster than vacancies, the self-interstitial flux wins over the vacancy flux. During crystal cooling, V-I recombination virtually eliminates the vacancies and the surviving self-interstitials are driven into supersaturation and finally aggregate. On the other hand, at fast pull rates the convective flux dominates and, due to the larger equilibrium concentration of vacancies compared to that of

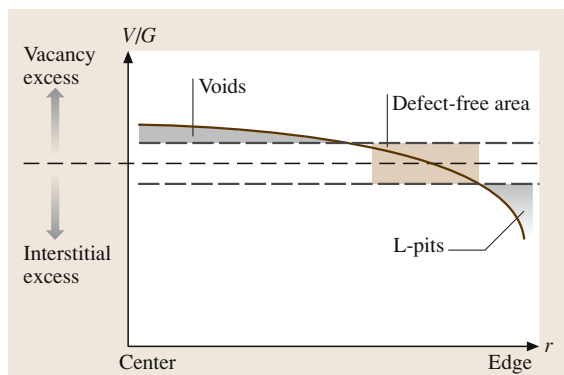


Fig. 5.9 Schematic radial variation in defect types as a function of the radial variation of V/G

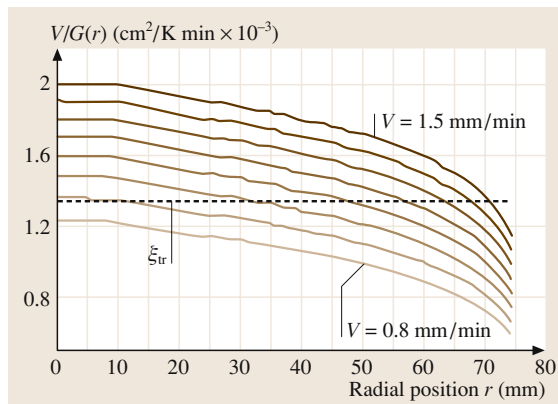


Fig. 5.10 Radial variation in the crossover between the transition value ξ_{tr} and $V/G(r)$ as function of decreasing pull rate V

self-interstitials, more vacancies flow into the crystal, resulting in vacancy aggregates. A similar change in the prevailing defect type is obtained when G is varied. The above model, originally proposed by Voronkov [5.62], yields the previously mentioned parameter V/G which determines the defect type.

In standard growth processes, G always exhibits a notable increase from the center towards the crystal rim, which (at sufficiently low pull rates) results in a vacancy excess in the center, while Si interstitials dominate at the crystal rim, as indicated in Fig. 5.9. Figure 5.10 illustrates how the radial position of the crossover between ξ_{tr} and $V/G(r)$ and hence the boundary between the void and L-pit regions is shifted when V is varied.

The exact value of ξ_{tr} is still under discussion, because G has to be taken directly at the interface where the thermal field in the crystal changes very rapidly [5.36, 74]. Therefore, as yet it has not been possible to accurately measure G at the growth interface [5.75]. Fortunately, computer simulations with sufficiently refined meshes permit a rather accurate calculation of G . Based on these calculations, a value of $0.13 \text{ mm}^2 \text{ min}^{-1} \text{ K}^{-1}$ has been determined from growth experiments, with hot zones of different G s [5.34], which is widely accepted for now.

It should be noted that the above behavior is probably unique to silicon as it stems from the peculiar phenomenon that the vacancy and Si interstitial equilibrium concentrations in crystalline silicon are very similar, but C_v^{eq} is slightly higher than C_I^{eq} , although the diffusivity of Si interstitials exceeds that of vacancies at the melting point. In addition, the quantitative relationship between

the relevant parameters which allows the changeover in defect behavior to occur in accessible growth rate regimes is quite astonishing. The time-dependent behaviors of the point defect concentrations are determined by the species conservation equations [5.76]

$$\begin{aligned} \frac{DC_I}{Dt} = & \nabla \left[D_I(T) \nabla C_I - \frac{D_I(T) C_I Q_I^*}{k_b T^2} \nabla T \right] \\ & - k_{IV} (C_I C_V - C_I^{\text{eq}} C_V^{\text{eq}}) \\ & - \frac{\partial}{\partial t} \int_2^\infty n f_1(n, r) dn, \end{aligned} \quad (5.1)$$

where Q_I^* is the reduced heat of transport for Si interstitials, which describes the rate of material flow due to a temperature gradient. A similar expression applies for vacancies (just exchange the I and V notations). The substantial derivative $D/Dt = [\partial/\partial t + V(\partial/\partial z)]$ is defined here as the rate of change of species concentration at a point moving with the velocity (V) of the advancing growth interface. This term relates to the incorporation of species due to convection. The first term on the left-hand side accounts for Fickian diffusion and the second for thermodiffusion. The last term represents the loss of point defects into their aggregates of all possible sizes. This term can be neglected as long as the supersaturation of the surviving species has not increased to the point where cluster nucleation starts. A detailed analysis also reveals that thermodiffusion has no detectable impact. Hence, the simple picture drawn above – which only takes into account convection and Fickian diffusion – is quite reasonable. Based on these simplifications, the incorporated vacancy concentration can be calculated from a simple expression [5.74, 77].

$$C_V(r) \approx [C_V^{\text{eq}}(T_m) - C_I^{\text{eq}}(T_m)] \left(1 - \frac{\xi_{tr} G(r)}{V} \right). \quad (5.2)$$

This expression is also applicable to Si interstitials (at $V/G < \xi_{tr}$). It then defines the incorporated Si interstitial concentration C_I with a minus sign.

If V/G approaches ξ_{tr} , then the vacancy and Si interstitial fluxes become equal and both species are annihilated by V–I recombination. Thus, the incorporated V and I concentrations are negligible and no aggregates can form. This particular condition must be met for the growth of so-called perfect silicon.

5.3.3 Theoretical Model: Aggregation of Intrinsic Point Defects

Due to the technological dominance of vacancy-rich silicon, previous theoretical investigations of defect aggregation have mainly focused on the formation of voids. In addition, it was also easier to verify the theoretical results for vacancy aggregation because voids are not obscured by the secondary defect generation that occurs for L-pits.

From Fig. 5.11, it can be seen that, for a standard growth process, the V-supersaturation already begins to build up at around 1200 °C [5.78], which is well above the experimentally determined nucleation temperature of voids (≈ 1100 °C). Thus, the supersaturation C_V/C_V^{eq} necessary to nucleate voids is appreciable, ≈ 10 [5.76], as estimated from Fig. 5.11. This fact also allows us to decouple the phase where V–I recombination is the major vacancy loss mechanism from the phase where nucleation dominates the vacancy loss, so that one can assume that the cluster formation starts with a fixed concentration which can be calculated from (5.2). For special growth processes that are designed to minimize or even eliminate voids, the starting vacancy concentration after V–I recombination is so small that the beginning of supersaturation (and hence void nucleation) occurs at notably lower temperatures [5.77].

The first vacancy aggregation calculations were performed with a model that was originally developed to simulate oxygen precipitation [5.59, 79]. For the formation of small clusters (less than 20 vacancies), the kinetics are described by chemical rate equations; for

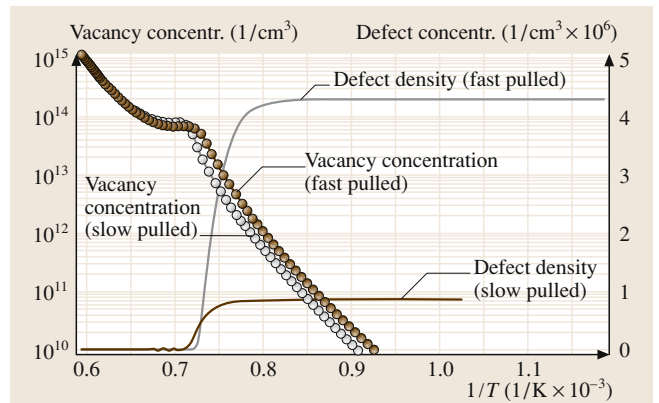


Fig. 5.11 Simulated variation in the vacancy concentration in growing crystals as a function of temperature for rapidly and slowly cooling growth processes. The density of voids nucleated at around 1100 °C and below is shown on the right scale

larger clusters, the rate equations are expanded into a continuum formulation that yields a single Fokker–Planck equation. The void shape is assumed to be spherical. Input data are the computed temperature field, which can be quite accurately calculated using commercially available standard codes, and the surviving vacancy concentration after V–I recombination has ceased. Besides the physical properties of the vacancies and Si interstitials, the surface energy must also be known, which is estimated to be around 950 erg/cm^2 .

As can be seen from Fig. 5.11, the computer simulations yield a nucleation temperature of around 1100°C , which agrees fairly well with the experimental results. The prediction of the correct aggregation temperature has been a matter of discussion. With the above input data, one ends up with a far lower aggregation temperature than those observed [5.80, 81]. The main reason for this is that the nucleation process does not start as a small octahedral void as previously assumed, but rather as an “amorphous cloud” of vacancies which, after some time, relaxes into an octahedron with (111) facets [5.82]. Proper modeling of this initial nucleation process with atomistic calculations does indeed shift the calculated aggregation temperature into the experimentally observed range.

The results in Fig. 5.11 also demonstrate that a higher cooling rate obviously decreases the nucleation temperature. The origin of this effect is the increased supersaturation, which builds up because there is less time for vacancy consumption. Due to the shift in the nucleation temperature, the residual vacancy concentra-

tion after void formation remains higher for fast cooling rates than for slow cooling. This has an important effect on the nucleation of oxygen precipitates, as the latter is strongly enhanced by a higher free vacancy concentration. Figure 5.12 illustrates the evolution of the void density/size distribution during the nucleation period. It is apparent that the maximum of the distribution is shifted to larger void sizes not only due to the growth of the previously nucleated voids, but also due to the fact that the newly nucleated clusters have a larger size. The latter effect results from the decreased supersaturation with higher void density, which, in turn, increases the critical radius for stable nuclei. Thus, only increasingly large clusters can nucleate towards the end of the nucleation period.

As depicted in Fig. 5.13, a fast cooling rate decreases the cluster size but increases their density and vice versa, which correlates well with the experimental data. The reason for this is the above-mentioned higher supersaturation at fast cooling rates, which entails a larger nucleation rate. The higher cluster density then effectively lowers the residual vacancy concentration. The clusters, however, have no time to grow and so remain small.

A similar effect is obtained when the initial vacancy concentration before the onset of nucleation is reduced [5.83]. A lower vacancy concentration causes a downward shift in the nucleation temperature, where the vacancy diffusivity is lower. As a consequence, they need more time to diffuse to sinks, which in turn drives up supersaturation and so more clusters of a smaller size are formed again. The behavior of the size distribution as a function of the initial vacancy concentration gives a simple explanation for the experimentally found radial variation in the void size distribution in Fig. 5.7. As

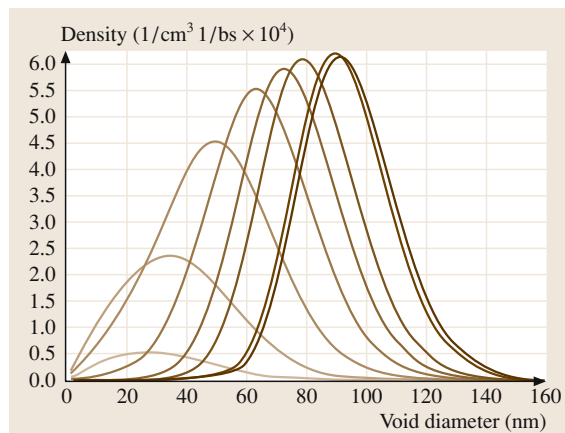


Fig. 5.12 Simulated evolution of the void density–size distribution at various temperatures of the growing crystal. The temperature difference between each curve is 4 K

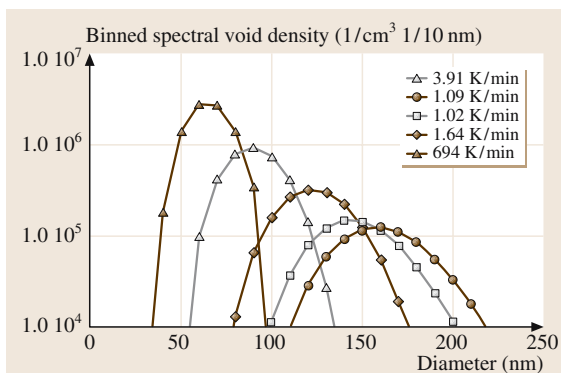


Fig. 5.13 Simulated density/size distributions of voids for growth processes with different cooling rates

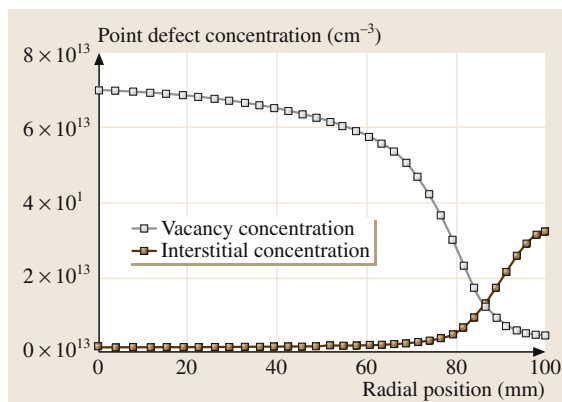


Fig. 5.14 Simulated radial variations in the vacancy and Si interstitial concentrations, respectively. (After [5.68])

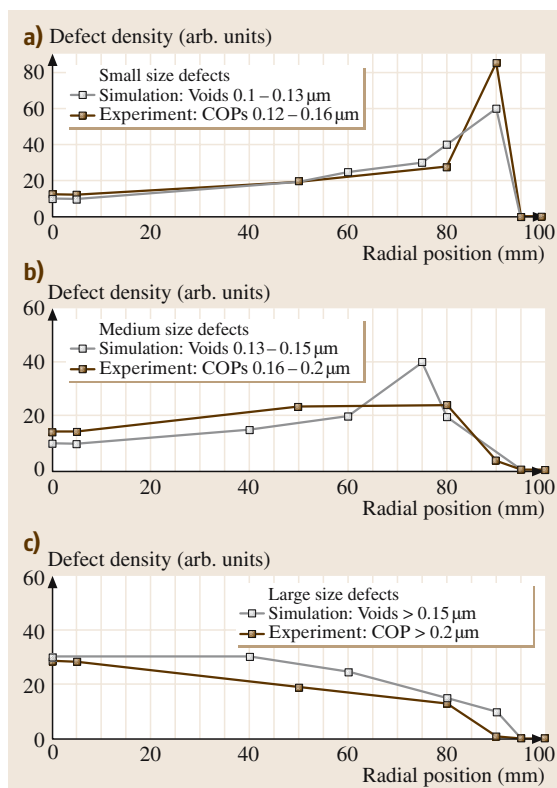


Fig. 5.15 Simulated radial variation in the density/size distribution of voids and comparison with experimental data. (After [5.68])

the initial vacancy concentration decreases towards the crystal rim (Fig. 5.14), the average size of a void shrinks but void density grows. Thus, the density of small voids

is low in the center, but increases towards the crystal rim until the vacancy concentration drops below a critical value, where the average void size is below the detection limit. On the other hand, the number of large voids shrinks with decreasing vacancy concentration. Therefore, the density of large COPs is highest in the center. This behavior is also fairly well reproduced by simulation (Fig. 5.15).

A simple expression that predicts a void density proportional to $q^{3/2}C_V^{-1/2}$ (q = cooling rate, C_V = initial vacancy concentration before nucleation starts) has been derived by Voronkov et al. [5.66]. This formula is obviously in excellent agreement with the experimental results (Fig. 5.16). A deviation from the above expression was reported for detached crystals, where significantly higher cooling rates can be achieved (2–70 K/min) [5.84]. In the latter case, this discrepancy was tentatively explained by a suppression of voids in favor of enhanced formation of oxygen precipitates at high cooling rates (> 40 K/min) [5.85]; in other words, so-called “oxide particles” nucleate at around 1100 °C rather than voids.

The $C^{-1/2}$ dependence is more difficult to verify experimentally because a change in the initial vacancy concentration entails a change in the nucleation temperature and so temperature-dependent parameters like diffusivities and equilibrium concentrations must be adjusted. Most of the current aggregation models do not consider the oxidation of the inner void surface, which stops void growth before the vacancies are essentially depleted. This has less impact on the void size distribution but considerably affects the residual vacancy concentration and therefore oxygen precipitation (Sect. 5.4).

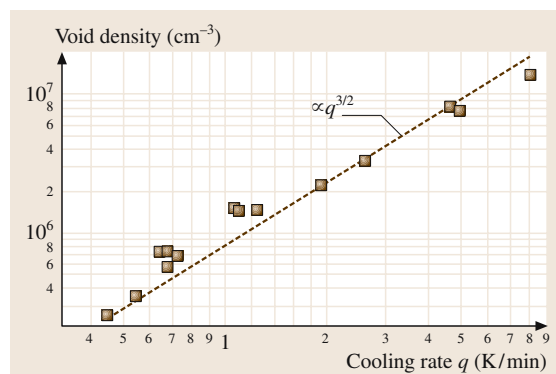


Fig. 5.16 Void density as derived from GOI measurements as a function of cooling rate of the growing crystal. The straight line relates to predictions of the $q^{3/2}$ law

The occurrence of double voids is still under discussion. One attempt to explain this phenomenon is based on a partial oxidation of the (111) facets during void growth, which allows the incorporation of further vacancies at the unoxidized residual facet surface [5.86].

5.3.4 Effect of Impurities on Intrinsic Point Defect Aggregation

Nitrogen

Nitrogen has long been known to simultaneously suppress interstitial- and vacancy-related defects – A-swirls and D-defects/voids, respectively [5.88] – in floating zone (FZ)-grown crystals. The nitrogen concentrations required are very low ($< 2 \times 10^{14}$ at/cm³) and were found to be in the same range as the surviving residual intrinsic point defect concentrations.

In the case of Czochralski (CZ)-grown silicon, nitrogen has been an undesirable dopant in the past, because it enhances the formation of oxidation-induced stacking faults (OSF) [5.89]. As the nitrogen concentration increases, the OSF ring widens [5.90] until it extends over the entire crystal diameter. In addition, due to the comparatively high oxygen concentration ($3\text{--}8 \times 10^{17}$ at/cm³), the desirable defect suppressing effects of nitrogen are largely offset by the interaction of nitrogen and oxygen [5.91]. However, nitrogen still has a notable effect on the defect size: the higher the nitrogen concentration, the lower the defect size [5.19, 92]. This effect has recently become important, because a small defect size is highly favorable for defect annealing [5.16]. It was further found that nitrogen not only

reduces the defect size but also systematically changes the void morphology from octahedron to platelet and, finally, to rods or even rod clusters with increasing concentration (Fig. 5.17) [5.87]. Nitrogen doping has gained additional interest as it also strongly enhances oxygen precipitation [5.93–95] which can be used to improve the efficiency of gettering of metallic contaminants in low thermal budget device processes [5.96, 97].

Unfortunately, quantitative data on the chemical and physical properties of nitrogen in silicon are rare. Experimentally, it was found that nitrogen exists as N–N dimers in as-grown silicon [5.98], which has a diffusivity three orders of magnitude higher than oxygen at 1100 °C [5.99]. These diffusion experiments indicate that the N–N dimer is stable up to at least 1270 °C [5.100], which is in line with first-principles calculations of the N–N dimer configuration [5.101–103]. In contrast with the nitrogen molecule, the N–N dimer in silicon does not have any direct bonds between the two neighboring nitrogen atoms. Experiments with nitrogen-implanted silicon samples heated to melt temperature by a laser pulse and subsequently quenched to room temperature proved that roughly 10% of the nitrogen can be frozen on substitutional sites as single nitrogen atoms [5.104, 105]. The fact that substitutional nitrogen is generally not observed in regularly cooled samples suggests that substitutional nitrogen is not stable in silicon at lower temperatures [5.106].

As the nitrogen concentration needed to suppress A-swirl and D-defects is in the same range as the intrinsic point defect concentration before aggregation, it was suggested that nitrogen directly interacts with vacancies (V) and Si interstitials (I) and thus prevents their aggregation [5.107, 108]. One model proposes that the storage of vacancies in impurity complexes would gradually shift the transition value ξ_{tr} of V/G to higher values which, in turn, would explain the shrinkage of the void region with increasing nitrogen content [5.109]. On the other hand, the model also seems to imply that the A-swirl region simultaneously expands toward the crystal center, which is in conflict with experimental results. In particular, the Li-drift experiments of Knowlton et al. [5.110] proved that, despite the suppression of detectable COPs, the extent of the vacancy-rich region does not change due to nitrogen doping. Other attempts are based on the reduction of vacancy diffusivity due to the vacancy–nitrogen interaction [5.111] or on increasing the reaction barrier to vacancy absorption at the void interface [5.112]. However, the fact that nitrogen doping has no impact on the inner oxide layer growth of the voids [5.87] does not support the notion that nitro-

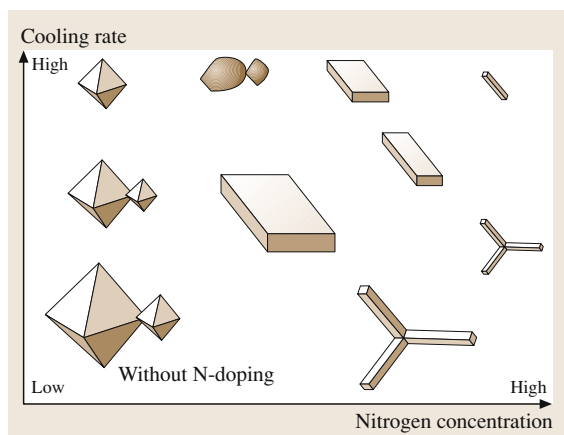
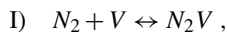


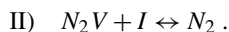
Fig. 5.17 Variation in void morphology as a function of nitrogen concentration and cooling rate of the growing crystal. (After [5.87])

gen modifies the properties of the void interface. None of these proposals gives any explanation for influence of nitrogen on Si interstitial aggregation.

First-principles calculations by Sawada et al. and Kageshima et al. [5.101, 102] have shown that the N–N dimer can form stable N_2V and N_2V_2 complexes, the latter being more stable due to a larger enthalpy of formation. Based on these results, it has been proposed that the vacancy aggregation is suppressed by the reaction

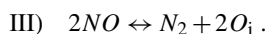


whereas the formation of Si interstitial agglomerates is prevented by the reaction



Nitrogen is not likely to be incorporated as a N–N dimer into the crystal. This is concluded from the experimental fact that nitrogen does not evaporate from the silicon melt and so it must be strongly bound to silicon atoms, which is not possible for a nitrogen molecule. It was therefore proposed that single nitrogen atoms are incorporated on interstitial (N_i) as well as substitutional sites (N_s) at the growth interface [5.113]. Behind the growth interface, the highly mobile interstitial nitrogen atoms then quickly form N–N dimers via $N_i + N_i \leftrightarrow N_2$ and nitrogen-vacancy complexes via $N_i + N_s \leftrightarrow N_2V$ inside the crystal. A ratio of $[N_i]:[N_s]=7:1$, which corresponds to 12% of the nitrogen at substitutional sites, has been determined from FZ crystals grown with various nitrogen contents. This is in good agreement with the above-mentioned experimental findings for substitutional nitrogen. As vacancies and Si interstitials have very similar concentrations at the melting point, reactions (I) and (II) mainly act as an additional V–I recombination path during crystal cooling, with only marginal impact on the levels of N_2 and N_2V . However, when recombination is complete and the vacancies or the Si interstitials have been wiped out, reaction (I) or (II) prevents the aggregation of the surviving species. Using this model, it has been possible to determine the radial variations of vacancies and Si interstitials in FZ crystals before the onset of defect aggregation [5.113]. This was achieved by measuring the radial extension of the COP and A-swirl region as a function of the nitrogen concentration. The result is in excellent agreement with theoretical predictions (Fig. 5.18). Surprisingly, the experimental data could not be fitted by assuming that the formation of N_2V_2 complexes prevails, as predicted by theoretical calculations. It is not clear whether the reaction to N_2V_2 is hampered by a high-energy barrier.

While this model works well for FZ crystals, it does not explain why the effect of nitrogen on defect aggregation is strongly diminished in CZ crystals. It is clear that the well-known interaction between nitrogen and oxygen, which results in the formation of N_xO_y complexes [5.114], is probably the origin of the vanishing nitrogen effect. However, it is also known that, despite oxygen levels of $5\text{--}6 \times 10^{17} \text{ at/cm}^3$, ample unreacted nitrogen is left in oxygen- and nitrogen-doped crystals to suppress defect aggregation [5.115]. As the experimentally confirmed N_2O complex is only stable up to about 700°C [5.114, 116], it has been proposed [5.113] that nitrogen reacts with oxygen at high temperatures according to



The existence of the NO complex has been theoretically predicted by Gali et al. [5.117]. It has a similar structure to the N_2 complex but with a smaller binding energy. The NO formation can occur directly close to the growth interface via $N_i + O_i \leftrightarrow NO$ or $N_s + O_i \leftrightarrow NO + V$. The latter reaction obviously generates additional vacancies, which should, however, only have a small impact on the total vacancy concentration because the close proximity of the growth interface and the high diffusivity of the vacancies will not allow large deviations from the equilibrium value. Due to the large excess of oxygen over nitrogen, it has been assumed that the equilibrium of reaction (III) is on the left-hand side near the melting point and so very little N_2 is available. As the temperature decreases, the

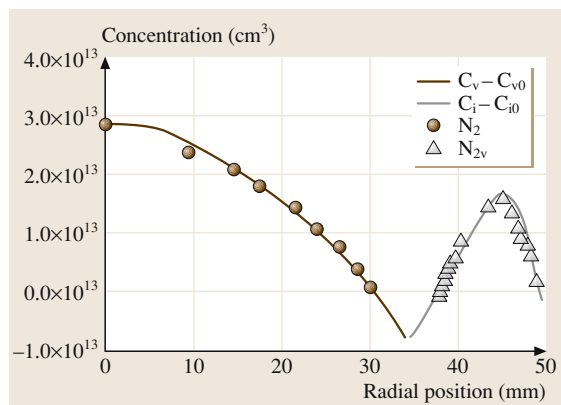


Fig. 5.18 Radial vacancy and Si interstitial distributions as derived from a FZ crystal grown with an axially varying nitrogen concentration. $C_{V0} = 8.8 \times 10^{12} \text{ at/cm}^3$ and $C_{I0} = 7.7 \times 10^{12} \text{ at/cm}^3$. The solid and dashed lines refer to simulation results. (After [5.113])

equilibrium gradually shifts to the right-hand side and more and more N_2 is generated. If the temperature at which N_2 starts to develop in sufficient quantities is below the vacancy aggregation temperature, the formation of voids is not suppressed. If the N_2 formation temperature is above the vacancy aggregation temperature, void formation depends on how much N_2 is available in relation to the vacancy concentration at the void nucleation temperature. The higher the N_2 formation temperature, the more N_2 is available to reduce the number of vacancies by reaction (I). According to reaction (III), the formation temperature of N_2 should increase with lower oxygen and higher nitrogen concentrations. Lower free vacancy concentrations and therefore lower aggregation temperatures result in voids of smaller size but higher density, in agreement with the experimental findings.

After void nucleation, the free vacancy concentration drops rapidly and the above reaction (I) reverses, giving $N_2V \rightarrow N_2 + V$, which slows down the decay of free vacancies. Thus, the N_2V -complex acts as an inter-reservoir for vacancies – its formation before void formation removes vacancies and its dissociation after void formation releases them again [5.90]. In conjunction with the lower vacancy aggregation temperature, this results in a notable increase in free vacancy concentration during this cool-down phase compared to nitrogen-free material (Fig. 5.19). As free vacancies are known to strongly favor the nucleation of oxygen precipitates, the latter effect gives a simple explanation for the observed strong enhancement of oxygen precipitation in nitrogen-doped CZ crystals (Sect. 5.4).

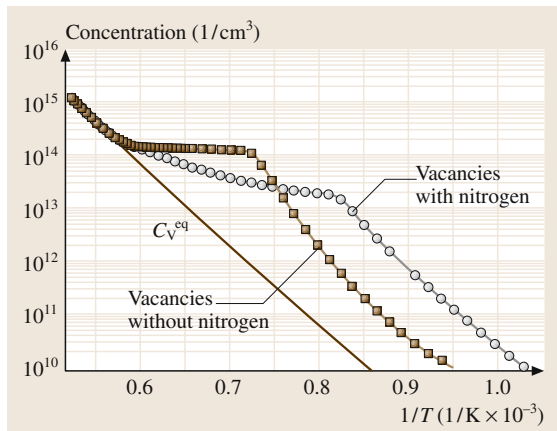


Fig. 5.19 Simulated variations in the vacancy concentration as a function of temperature for a nitrogen-doped and an undoped crystal

The remarkable variation in the void morphology as a function of the nitrogen content is still under discussion. A similar change in morphology is also known for oxygen precipitates. In this case, the effect can be interpreted in terms of the balance between the relaxation energy of the lattice strain and the interfacial energy. However, TEM investigations do not reveal any lattice strain around voids, and so another mechanism is required to explain the varying void morphology.

Boron

Boron doping affects the intrinsic point defect aggregation in a similar way to nitrogen doping, but much higher concentrations are required [5.119]. It was found that the COP region as well as the diameter of the OSF ring starts to shrink when the boron content exceeds roughly 5×10^{18} at/cm³ (Fig. 5.20). In contrast with nitrogen, OSF formation is not enhanced and the width of the OSF ring remains unchanged with increasing boron concentration. At around 10^{19} at/cm³, the COPs and the OSF ring disappear in the center of the crystal. Although detailed investigations into the behavior of L-pits in highly boron-doped silicon are still required, we can infer from the defect-free quality of pp+ epi wafers that L-pits are simultaneously suppressed by the high boron level (Sect. 5.3.1).

Dornberger et al. [5.67] suspected that changes in the equilibrium concentrations and diffusivities of intrinsic point defects in the presence of high boron concentrations would modify the transition value ξ_{tr} of the parameter V/G as a function of the boron content. Voronkov and Falster [5.81] also proposed a modification of equilibrium concentrations as a result of the shift

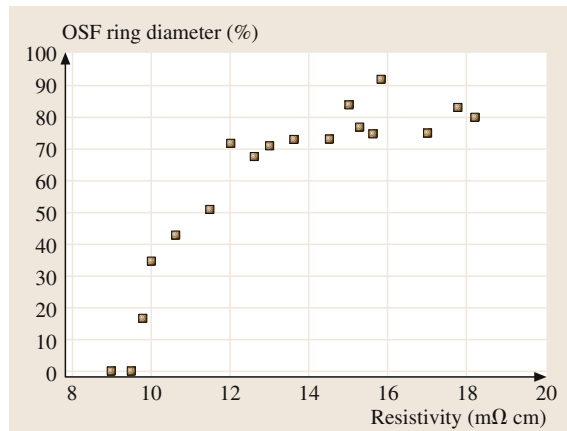


Fig. 5.20 Variation in the OSF ring diameter as a function of the boron concentration. (After [5.118])

in the Fermi level due to the high boron concentration. Based on this approach, the latter authors calculated a shift in the transition value ξ_{tr} which would account for the observed shrinkage of the void region. As with nitrogen doping, the ξ_{tr} shift entails the simultaneous appearance of L-pits in the outer crystal region and a shrinking void region, which is in conflict with experimental results. Another attempt by Sinno et al. [5.120] considers reversible reactions between boron and intrinsic point defects, in particular self-interstitials. The formation of BI and B₂I complexes consumes self-interstitials which would otherwise be annihilated by recombination. As recombination consumes the remaining self-interstitials, BI and B₂I start to dissociate again according to Le Chatelier's principle. As a result, the self-interstitial concentration increases, which shifts the point defect balance towards the interstitial-rich side, causing the void region to shrink. No comment is made about whether or not the model is also able to account for the suppression of L-pit formation by storing supersaturated self-interstitials in boron complexes.

Carbon

Carbon doping is also known to impact intrinsic point defect aggregation. It has been reported that the void region in the crystal center disappears upon carbon doping [5.57, 121], while the region of Si interstitial aggregates is widened; in other words, carbon does not inhibit the formation of L-pits/A-swirl. This behavior appears to be in line with the model put forward

in [5.109], which predicts an upward shift of ξ_{tr} for carbon doping and, in turn, an enlargement of the interstitial-rich region.

It was also found that carbon doping reduces the grown-in defect size [5.87, 122]. Although the size reduction is appreciable, the morphology of the voids is not changed, in contrast to the consequences of nitrogen doping. Only a tendency towards multiple void formation was observed. The thickness of the inner oxide layer was found to be similar to that of undoped crystals, which indicates that carbon doping has no influence on the growth of the inner oxide layer. As with boron, the effect on defect aggregation is only seen at significantly higher concentrations ($\approx 1 \times 10^{17}$ at/cm³) than for nitrogen.

At higher concentrations, carbon is known to enhance oxygen precipitation too [5.123, 124]. As carbon predominantly resides on substitutional sites, it is very unlikely that, as in the case of nitrogen doping, a higher residual vacancy concentration is responsible for the stronger oxygen precipitation. On the other hand, the small carbon atom exerts a local tensile strain on the surrounding lattice and attracts Si interstitials to form C_xI_y complexes. Thus, Si interstitials ejected into the lattice by growing oxygen precipitates are effectively removed and, in turn, further precipitate growth is not retarded by a build-up of Si interstitial supersaturation. The enhanced oxygen precipitation may also be related to heterogeneous nucleation at small carbon aggregates [5.125].

5.4 Formation of OSF Ring

The ring-like distributed OSFs are oxygen precipitates with platelet shapes that grow particularly large at the edge of the void region, and exceed a critical size necessary to create stacking faults during subsequent wafer oxidation there [5.126]. The critical radius of the grown-in platelets is ≈ 70 nm. The formation of OSFs can be suppressed if the cooling rate of the growing crystal is increased (this means, for example, that OSFs are not found in oxygen-doped FZ crystals [5.118] which have very high cooling rates) or if the oxygen content is decreased.

The peculiar ring-like distribution is a consequence of the well-known strong enhancement of oxygen precipitation by vacancies [5.39, 127]. The reason for this is that the absorption of vacancies allows the oxygen precipitate, which occupies twice as much volume as the corresponding silicon lattice, to nucleate and grow

without building up notable strain energy. As will be described below, the radially inhomogeneous oxygen precipitation is a consequence of a substantial radial inhomogeneity of the free vacancy concentration in contrast to the rather flat radial profile of oxygen.

The vacancy concentration, as pointed out previously, has its maximum at the center of the growing crystal after V-I recombination has ceased (Fig. 5.21). Therefore, the critical supersaturation for void formation is first reached at the crystal center at relatively high temperatures. Hence, the free vacancies are quickly consumed in this area. As the crystal cools, voids are also nucleated in the regions of lower initial vacancy concentration, meaning that vacancy consumption then also occurs further away from the crystal center. As the removal of vacancies is enhanced at higher temperatures, where the diffusivity is large, the ra-

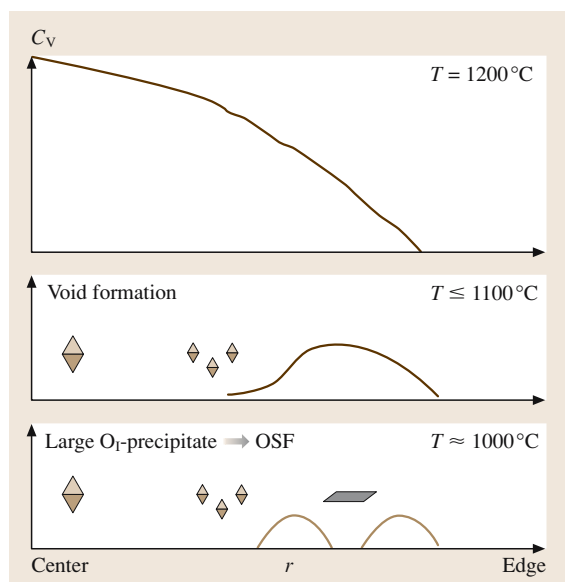


Fig. 5.21 Schematic evolution of the radial vacancy profile and accompanying void and large O_I -precipitate formation at decreasing temperatures

dial vacancy distribution finally drops to a minimum in the crystal center, and the maximum of the residual vacancies gradually moves toward the V/I boundary upon further cooling. Because of this shifted vacancy maximum, the first oxygen aggregates nucleate close to the V/I boundary on the vacancy-rich side. The relatively high nucleation temperature at this position results in larger sizes but lower densities for the corresponding grown-in oxygen precipitates compared to those nucleated at lower temperatures away from the V/I boundary. Consequently, they preferentially reach the critical size for OSF formation during subsequent

wafer oxidation. At sufficiently fast crystal cooling rates, the oxygen precipitates are nucleated at higher densities but smaller sizes and cannot grow to the critical size. Thus, OSF formation is prevented, in agreement with experimental findings. The same density/size effect is obtained if the oxygen content is lowered – similar to the behavior of voids – and so OSFs are also suppressed.

The nucleations of the first oxygen precipitates quickly results in the depletion of residual vacancies in their neighborhood. Thus, the previous vacancy maximum at the V–I boundary is then converted into a pronounced minimum which suppresses further nucleation in this area. As a result, two vacancy peaks with a region of large oxygen precipitates in-between is frozen in as the crystal cools down to room temperature (Fig. 5.21). Later on, when the wafers are subjected to further heat treatments, such as a nucleation step at 780°C followed by a growth step at 1000°C, new oxygen precipitates with high density are preferentially nucleated at the radial positions of the two remaining vacancy maxima. This results in a distinct profile for the interstitial oxygen concentration after these heat treatments. In the radial band where the large precipitates and the OSF ring are located, the interstitial oxygen concentration is only slightly reduced because further nucleation of precipitates during wafer heat treatment is suppressed there and the comparatively low density of precipitates does not consume much of the interstitial oxygen. On the other hand, the high density bands at the locations adjacent to the former vacancy peaks coincide with pronounced minima in the interstitial oxygen content.

Considerable changes in the widths and the radial positions of OSF rings are observed if the crystals are nitrogen-doped. With increasing nitrogen content, the outer and inner OSF ring boundaries are shifted towards the crystal center [5.90]. However, the shift is much larger for the inner OSF ring boundary, resulting in a significant widening of the OSF ring width (Fig. 5.22) until it extends over the entire crystal volume. This behavior was tentatively explained by a schematic model that takes into account the slower build-up of vacancy supersaturation due to the formation of N_2V complexes and the resulting lower void nucleation temperature. The model also predicts that the void and the OSF region overlap, roughly in the area between the inner boundary and the middle of the OSF ring, which is in agreement with experimental data (Fig. 5.23).

In contrast to nitrogen, carbon tends to suppress OSF formation. This is again an indication that the

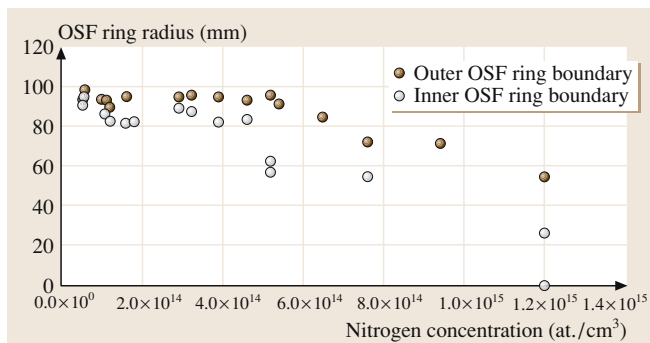
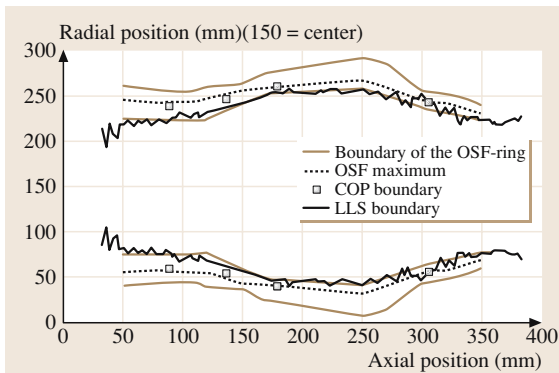


Fig. 5.22 Variations of the outer and inner OSF ring boundaries as a function of the nitrogen concentration. (After [5.90])



mechanism of the effect of carbon is different. The above-mentioned idea of heterogeneous nucleation due

Fig. 5.23 Variations in the OSF ring width and its radial location in comparison to the variations of the COP and LLS boundaries along the crystal axis. The LLS defects were detected on the as-polished wafer surface, which detects only the larger voids, while the COP boundary was determined by LLS measurements after repeated SC1 cleaning, which enlarges the voids to above the LLS detection limit

to small carbon clusters may also explain this behavior. As heterogeneous nucleation needs less supersaturation, it prevails over homogeneous nucleation. Furthermore, the density of carbon clusters is much higher than that of homogeneously nucleated oxygen precipitates. Thus, the heterogeneously nucleated oxygen precipitates cannot grow to the critical size due to their large densities.

References

- 5.1 P. M. Petroff, A. J. R. deKock: *J. Cryst. Growth* **30**, 117 (1975)
- 5.2 S. Sadamitsu, S. Umeno, Y. Koike, M. Hourai, S. Sumita, T. Shigematsu: *Jpn. J. Appl. Phys.* **32**, 3675 (1993)
- 5.3 L. I. Bernewitz, B. O. Kolbesen, K. R. Mayer, G. E. Schuh: *Appl. Phys. Lett.* **25**, 277 (1975)
- 5.4 B. O. Kolbesen, A. Mühlbauer: *Solid State Electron.* **25**, 759 (1982)
- 5.5 W. Bergholz, W. Mohr, W. Drewes: *Mater. Sci. Eng.* **B4**, 359 (1989)
- 5.6 M. Itsumi, H. Akiya, T. Ueki, M. Tomita, M. Yamawaki: *J. Appl. Phys.* **78**(10), 5984 (1995)
- 5.7 M. Miyazaki, S. Miyazaki, Y. Yanase, T. Ochiai, T. Shigematsu: *Jpn. J. Appl. Phys.* **34**, 6303 (1995)
- 5.8 J. G. Park, J. M. Park, K. C. Cho, G. S. Lee, H. K. Chung: Effect of crystal defects in device characteristics, Proc 2nd Int. Symp. on Advanced Science and Technology of Silicon Materials, Kona-Hawaii, USA 1996, ed. by M. Umeno (145th Committee of the Japan Society for the Promotion of Science, 1996) 2519
- 5.9 M. Itsumi: *Mater. Sci. Eng.* **B73**, 184 (2000)
- 5.10 T. Bearda, M. Houssa, P. Mertens, J. Vanhellemont, M. Heyns: *Appl. Phys. Lett.* **75**(9), 1255 (1999)
- 5.11 U. Lambert, A. Huber, J. Grabmeier, J. Vanhellemont, R. Wahlich, G. Kissinger: *Microelectron. Eng.* **48**, 127 (1999)
- 5.12 E. Dornberger, D. Temmler, W. v. Ammon: *J. Electrochem. Soc.* **149**(4), 1 (2002)
- 5.13 C. Kupfer, H. Roth, H. Dietrich: *Mat. Sci. Semicon. Proc.* **5**, 381 (2003)
- 5.14 A. M. Eidenzon, N. I. Puzanov: *Inorg. Mater.* **33**(3), 272 (1997)
- 5.15 J. G. Park, H. K. Chung: Wafer requirements: Memory devices, Silicon Wafer Symp., Portland 1999 (SEMICON West 99, Portland 1999) D-1
- 5.16 J. G. Park: *J. Jpn. Assoc. Crystal Growth* **27**(2), 14 (2000)
- 5.17 R. Falster: Advances of the defect engineering of polished silicon wafers: Perfect silicon and magic denuded zones, Silicon Wafer Symp., Portland 1999 (SEMICON West 99, Portland 1999) E-13
- 5.18 D. Gräf, M. Suhren, U. Lambert, R. Schmolke, A. Ehlert, W. v. Ammon, P. Wagner: *J. Electrochem. Soc.* **145**(1), 275 (1998)
- 5.19 D. Gräf, M. Suhren, U. Lambert, R. Schmolke, A. Ehlert, W. v. Ammon, P. Wagner: *J. Electrochem. Soc.* **145**(1), 275 (1998)
- 5.20 X. Yu, D. Yang, X. Ma, L. Li, D. Que: *Semicond. Sci. Technol.* **18**, 399 (2003)
- 5.21 K. Sumino, I. Yonenaga, M. Imai, T. Abe: *J. Appl. Phys.* **54**(9), 5016 (1983)
- 5.22 L. Jastrzebski, G. W. Cullen, R. Soydan, G. Harbeke, J. Lagowski, S. Vecrumba, W. N. Henry: *J. Electrochem. Soc.* **134**(2), 466 (1987)
- 5.23 G. Wang, D. Yang, D. Li, Q. Shui, J. Yang, D. Que: *Physica B* **308-310**, 450 (2001)
- 5.24 D. Maroudas, R. Brown: *Appl. Phys. Lett.* **62**(2), 172 (1993)
- 5.25 A. Seeger, K. P. Chik: *Phys. Status Solidi A* **29**, 455 (1968)
- 5.26 H. R. Schober: *Phys. Rev. B* **39**, 13013 (1989)
- 5.27 R. Car, P. J. Kelly, A. Oshiyama, S. T. Pantelides: *Phys. Rev. Lett.* **52**, 1814 (1984)
- 5.28 D. Maroudas, R. Brown: *Phys. Rev. B* **47**(23), 15562 (1993)
- 5.29 G. D. Watkins: *Mater. Sci. Semicond. Process.* **3**, 227 (2000)

- 5.30 N.A. Stolwijk, J. Holzl, W. Frank, E.R. Weber, H. Mehrer: Appl. Phys. A **39**, 37 (1986)
- 5.31 H. Bracht, N.A. Stolwijk, H. Mehrer: Phys. Rev. B **52**, 16542 (1995)
- 5.32 H. Zimmermann, H. Ryssel: Appl. Phys. A **55**, 121 (1992)
- 5.33 H. Bracht: Native point defects in silicon, Proc. 3rd Int. Symp. on Defects in Silicon III, Seattle, Washington 1999, ed. by T. Abe, W.M. Bullis, S. Kobayashi, W. Lin, P. Wagner (The Electrochemical Society, Pennington 1999) 357
- 5.34 W.v. Ammon, E. Dornberger, H. Oelkrug, H. Weidner: J. Cryst. Growth **151**, 273 (1995)
- 5.35 M. Hourai, E. Kajita, T. Nagashima, H. Fujiwara, S. Umeno, S. Sadamitsu, S. Miki, T. Shigematsu: Mater. Sci. Forum **196–201**, 1713 (1995)
- 5.36 E. Dornberger, W.v. Ammon: J. Electrochem. Soc. **143**(5), 1648 (1996)
- 5.37 T. Sinno, R.A. Brown, W.v. Ammon, E. Dornberger: Appl. Phys. Lett. **70**(17), 2250 (1997)
- 5.38 M. Akatsuka, M. Okui, N. Morimoto, K. Sueoka: Jpn. J. Appl. Phys. **40**, 3055 (2001)
- 5.39 R. Falster, V.V. Voronkov, F. Quast: Phys. Status Solidi B **222**, 219 (2000)
- 5.40 N. Fukata, A. Kasuya, M. Suezawa: Jpn. J. Appl. Phys. **40**, L854 (2001)
- 5.41 T. Frewen, T. Sinno, E. Dornberger, R. Hoelzl, W.v. Ammon, H. Bracht: J. Electrochem. Soc. **150**(11), G673 (2003)
- 5.42 D.A. Antoniadis, I. Moskowitz: J. Appl. Phys. **53**(10), 6780 (1982)
- 5.43 H.J. Gossmann, C.S. Rafferty, A.M. Vredenberg, H.S. Luftman, F.C. Unterwald, D.J. Eaglesham, D.C. Jacobson, T. Boone, J.M. Poate: Appl. Phys. Lett. **64**(3), 312 (1994)
- 5.44 T. Sinno: Thermophysical properties of intrinsic point defects in crystalline silicon, Proc 9th Int. Symp. on Silicon Materials Science and Technology, Semiconductor Silicon, Philadelphia 2002, ed. by H.R. Huff, L. Fabry, S. Kishino (The Electrochemical Society, Pennington 2002) 212
- 5.45 T. Ueki, M. Itsumi, T. Takeda: Jpn. J. Appl. Phys. **37**, 1669 (1998)
- 5.46 M. Itsumi: J. Cryst. Growth **237–239**, 1773 (2002)
- 5.47 S. Umeno, Y. Yanase, M. Hourai, M. Sano, Y. Shida, H. Tsuya: Jpn. J. Appl. Phys. **38**, 5725 (1999)
- 5.48 M. Nishimura, Y. Yamaguchi, K. Nakamura, J. Jablonski, M. Watanabe: Electrochem. Soc. Symp. Proc. **98–13**, 188 (1998)
- 5.49 J. Ryuta, E. Morita, T. Tanaka, Y. Shimanuki: Jpn. Appl. Phys. **29**, L1947 (1990)
- 5.50 H. Yamagishi, I. Fusegawa, N. Fujimaki, M. Katayama: Semicond. Sci. Techn. **7**, A135 (1992)
- 5.51 P.J. Roksnor, M.M.B. Van de Boom: J. Cryst. Growth **53**, 563 (1981)
- 5.52 H. Bender, J. Vanhellemont, R. Schmolke: Jpn. J. Appl. Phys. **36**, L1217 (1997)
- 5.53 R. Schmolke, W. Angelberger, W.v. Ammon, H. Bender: Solid State Phenom. **82–84**, 231 (2002)
- 5.54 K. Nakai, M. Hasebe, K. Ohta, W. Ohashi: J. Cryst. Growth **210**, 20 (2000)
- 5.55 H. Föll, B.O. Kolbesen: Appl. Phys. **8**, 319 (1975)
- 5.56 P.M. Petroff, A.J.R. de Kock: J. Cryst. Growth **36**, 1822 (1976)
- 5.57 J. Chikawa, T. Abe, H. Harada: Impurity effect on the formation of microdefects during silicon crystal growth. In: *Semiconductor Silicon*, ed. by H.R. Huff, T. Abe, B. Kolbesen (The Electrochemical Society, Pennington 1986) p. 61
- 5.58 H. Föll, U. Gösele, B.O. Kolbesen: J. Cryst. Growth **40**, 90 (1977)
- 5.59 E. Dornberger, J. Esfandyari, D. Gräf, J. Vanhellemont, U. Lambert, F. Dupret, W.v. Ammon: Simulation of grown-in voids in Czochralski silicon crystals, Crystalline Defects and Contamination Control: Their Impact and Control in Device Manufacturing II, Nürnberg 1997, ed. by B.O. Kolbesen, P. Stallhofer, C. Claeys, F. Tardiff (The Electrochemical Society, Pennington 1997) 40
- 5.60 R. Schmolke, M. Blietz, R. Schauer, D. Zemke, H. Oelkrug, W.v. Ammon, U. Lambert, D. Gräf: Advanced silicon wafers for 0.18 μm design rule and beyond: Epi and fLASH!, High Purity Silicon VI, Phoenix 2000, ed. by C.L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H.J. Dawson (The Electrochemical Society, Pennington 2000) 1
- 5.61 W.v. Ammon, E. Dornberger, P.O. Hansson: J. Cryst. Growth **198/199**, 390 (1999)
- 5.62 V.V. Voronkov: J. Cryst. Growth **59**, 625 (1982)
- 5.63 M. Hasebe, Y. Takeoka, S. Shinoyama, S. Naito: Ring-like distributed stacking faults in CZ-Si wafers. In: *Defect Control in Semiconductors*, ed. by K. Sumino (Elsevier, Amsterdam 1990) p. 157
- 5.64 H. Yamagishi, I. Fusegawa, K. Takano, E. Iino, N. Fujimaki, T. Ohta, M. Sakurada: Evaluation of FDPs and COPs in silicon single-crystals, Semiconductor Silicon, San Francisco 1994, ed. by H.R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994) 124
- 5.65 W.v. Ammon, E. Dornberger: *Properties of Crystalline Silicon*, EMIS Datareviews Series 20, ed. by R. Hull (INSPEC, London 1999) p. 37
- 5.66 V.V. Voronkov, R. Falster: J. Cryst. Growth **194**, 76 (1998)
- 5.67 E. Dornberger, D. Gräf, M. Suhren, U. Lambert, P. Wagner, F. Dupret, W.v. Ammon: J. Cryst. Growth **180**, 343 (1997)
- 5.68 E. Dornberger, J. Esfandyari, J. Vanhellemont, D. Gräf, U. Lambert, F. Dupret, W.v. Ammon: Simulation of non-uniform grown-in void distributions in Czochralski crystal growth, Semiconductor Silicon, San Francisco 1998, ed. by H.R. Huff, U. Gösele, H. Tsuya (The Electrochemical Society, Pennington 1998) 490
- 5.69 M. Hourai, T. Nagashima, E. Kajita, S. Miki: Oxygen precipitation behavior in silicon during Czochralski crystal growth, Semiconductor Silicon, San Francisco

- 1994, ed. by H. R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994) 156
- 5.70 T. Iwasaki, A. Tomiura, K. Nakai, H. Haga, K. Kojima, T. Nakashizu: Influence of cooling condition during crystal growth of CZ-Si on oxide breakdown property, *Semiconductor Silicon*, San Francisco 1994, ed. by H. R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994) 744
- 5.71 K. Takano, K. Kitagawa, E. Iino, M. Kimura, H. Yamagishi: *Mater. Sci. Forum* **196–201**, 1707 (1995)
- 5.72 M. Akatsuka, M. Okui, S. Umeno, K. Sueoka: *J. Electrochem. Soc.* **150**(9), G587 (2003)
- 5.73 J. Furukawa, H. Tanaka, Y. Nakada, N. Ono, H. Shiraki: *J. Cryst. Growth* **210**, 26 (2000)
- 5.74 V. V. Voronkov, R. Falster: *J. Appl. Phys.* **86**(11), 5975 (1999)
- 5.75 A. Natsume, N. Inoue, K. Tanahashi, A. Mori: *J. Cryst. Growth* **225**, 221 (2001)
- 5.76 T. Sinno, E. Dornberger, W. v. Ammon, R. A. Brown, F. Dupret: *Mater. Sci. Eng.* **28**, 149 (2000)
- 5.77 Z. Wang, R. Brown: *J. Crystal Growth* **231**, 442 (2001)
- 5.78 E. Dornberger, W. v. Ammon, D. Gräf, U. Lambert, A. Miller, H. Oelkrug, A. Ehlert: The impact of dwell time above 900 °C during crystal growth on the gate oxide integrity of silicon wafers, *Proc. 4th Int. Symp. on High Purity Silicon*, San Antonio 1996, ed. by C. L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H. J. Dawson (The Electrochemical Society, Pennington 1996) 140
- 5.79 J. Esfandyari, G. Hobler, S. Senkader, H. Pötzl, B. Murphy: *J. Electrochem. Soc.* **143**, 995 (1996)
- 5.80 V. V. Voronkov, R. Falster: *J. Cryst. Growth* **198/199**, 399 (1999)
- 5.81 V. V. Voronkov, R. Falster: *J. Appl. Phys.* **87**(9), 4126 (2000)
- 5.82 T. A. Frewen, S. S. Kapur, W. Haeckl, W. v. Ammon, T. Sinno: *J. Crystal Growth* **279**, 258 (2005)
- 5.83 N. I. Puzanov, A. M. Eidenzon: *Semicond. Sci. Technol.* **7**, 406 (1992)
- 5.84 K. Nakamura, T. Saishoji, J. Tomioka: *J. Cryst. Growth* **237–239**, 1678 (2002)
- 5.85 V. V. Voronkov: *Mater. Sci. Eng.* **B73**, 69 (2000)
- 5.86 V. V. Voronkov, R. Falster: *J. Crystal Growth* **226**, 192 (2001)
- 5.87 J. Takahashi, K. Nakai, K. Kawakami, Y. Inoue, H. Yokota, A. Tachikawa, A. Ikari, W. Ohashi: *Jpn. J. Appl. Phys.* **42**, 363 (2003)
- 5.88 T. Abe, M. Kimura: In: *Semiconductor Silicon, 1990*, ed. by H. R. Huff, K. Barraclough, J. Chikawa (The Electrochemical Society, Pennington 1990) p. 105
- 5.89 D.-R. Yang, Y.-W. Wang, H.-N. Yao, D.-L. Que: *Progress in Natural Science* **3**(2), 176 (1993)
- 5.90 W. v. Ammon, R. Hoelzl, T. Wetzel, D. Zemke, G. Raming, M. Blietz: *Microelectron. Eng.* **66**, 234 (2003)
- 5.91 W. v. Ammon, A. Ehlert, U. Lambert, D. Gräf, M. Brohl, P. Wagner: Gate oxide related bulk properties of oxygen doped floating zone and Czochralski silicon, *Semiconductor Silicon*, San Francisco 1994, ed. by H. R. Huff, W. Bergholz, K. Sumino (The Electrochemical Society, Pennington 1994) 136
- 5.92 K. Nakai, Y. Inoue, H. Yokota, A. Ikari, J. Takahashi, A. Tachikawa, K. Kitahara, Y. Ohta, W. Ohashi: *J. Appl. Phys.* **85**(8), 4301 (2001)
- 5.93 F. Shimura, R. S. Hockett: *Appl. Phys. Lett.* **48**, 224 (1986)
- 5.94 Q. Sun, K. H. Yao, H. C. Gatos, J. Lagowski: *J. Appl. Phys.* **71**(8), 3760 (1992)
- 5.95 K. Aihara, H. Takeno, Y. Hayamizu, M. Tamatsuka, T. Masui: *J. Appl. Phys.* **88**(6), 3705 (2000)
- 5.96 K. Nakai, Y. Inoue, H. Yokota, A. Ikari, J. Takahashi, W. Ohashi: Formation of grown-in defects in nitrogen doped CZ-Si crystals, *Proc 3rd Int. Symp. on Advanced Science and Technology of Silicon Materials*, Kona, Hawaii 2000, ed. by M. Umeno (145th Committee of the Japan Society for the Promotion of Science, Kona 2000) 88
- 5.97 D. Gräf, U. Lambert, R. Schmolke, R. Wahlich, W. Siebert, E. Daub, W. v. Ammon: 300 mm Epi pp-wafer: Is there sufficient gettering?, *Proc. 6th Int. Symp. on High Purity Silicon*, Seattle, Washington 2000, ed. by C. L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H. J. Dawson (The Electrochemical Society, Pennington 2000) 319
- 5.98 H. J. Stein: Nitrogen in crystalline silicon, *Proc. Int. Symp. on Oxygen, Carbon, Hydrogen and Nitrogen in Crystalline Silicon*, Boston 1986, ed. by J. C. Mikkelsen Jr., S. J. Pearton, J. W. Corbett, S. J. Pennycook (Materials Research Society, Pittsburgh 1986) 523
- 5.99 Y. Itoh, T. Abe: *Appl. Phys. Lett.* **53**(1), 39 (1988)
- 5.100 A. Hara, A. Ohsawa: Interaction of oxygen and other point defects in silicon crystals, *Proc. Int. Symp. on Advanced Science and Technology of Silicon Materials*, Kona, Hawaii 1991, ed. by K. Kohra (145th Committee of the Japan Society for the Promotion of Science, Kona 1991) 47
- 5.101 H. Sawada, K. Kawakami: *Phys. Rev. B* **62**(3), 1851 (2000)
- 5.102 H. Kageshima, A. Taguchi, K. Wada: *Appl. Phys. Lett.* **76**(25), 3718 (2000)
- 5.103 R. Jones, S. Öberg, F. B. Rasmussen, B. B. Nielson: *Phys. Rev. Lett.* **72**, 1882 (1994)
- 5.104 K. L. Brower: *Phys. Rev. B* **26**, 6040 (1982)
- 5.105 H. J. Stein: *Appl. Phys. Lett.* **47**(12), 1339 (1985)
- 5.106 K. Murakami, H. Itoh, K. Takita, K. Masuda: *Appl. Phys. Lett.* **45**(2), 176 (1984)
- 5.107 W. v. Ammon, D. Gräf, W. Zulehner, R. Schmolke, E. Dornberger, U. Lambert, J. Vanhellemont, W. Hensel: Suppression of point defect aggregation in FZ silicon single crystals by nitrogen doping; Extending Abstracts, *Semiconductor Silicon*, San Diego 1998, ed. by H. R. Huff, U. Gösele, H. Tsuya (The Electrochemical Society, Pennington 1998) Abstract no. 512

- 5.108 K. Nakamura, T. Saishoji, S. Togawa, J. Tomioka: The effect of nitrogen on the grown-in defect formation in CZ silicon crystals. In: *Proceedings of the Kazusa Akademia Park Forum on the Science and Technology of Silicon Materials*, ed. by K. Sumino (Kazusa Akademia Park, Chiba 1999) p. 116
- 5.109 V. V. Voronkov, R. Falster: *J. Electrochem. Soc.* **149**(3), G167 (2002)
- 5.110 W. B. Knowlton, J. T. Walton, J. S. Lee, Y. K. Wong, E. E. Haller, W. v. Ammon, W. Zulehner: *Mater. Sci. Forum* **196–201**, 1761 (1995)
- 5.111 T. Ono, S. Umeno, T. Tanaka, E. Asayama, M. Hourai: Behavior of defects in nitrogen doped CZ-Si crystals, *Proc. Int. Symp. of the Forum on the Science and Technology of Silicon Materials*, Shonan Village Center, Kanagawa 2001, ed. by H. Yamata-Kaneta, K. Sumino (Japan Technical Information Service, Tokyo 2001) 95
- 5.112 K. Nakamura, T. Saishoji, S. Togawa, J. Tomioka: Influence of nitrogen on the point defect reaction in silicon, *Proc. Int. Symp. of the Forum on the Science and Technology of Silicon Materials*, Shonan Village Center 2001, ed. by H. Yamata-Kaneta, K. Sumino (Japan Technical Information Service, Tokyo 2001) 109
- 5.113 W. v. Ammon, R. Hölzl, J. Virbulis, E. Dornberger, R. Schmolke, D. Gräf: *J. Cryst. Growth* **226**(1), 19 (2001)
- 5.114 P. Wagner, R. Oeder, W. Zulehner: *Appl. Phys. A* **46**, 73 (1988)
- 5.115 W. v. Ammon, P. Dreier, W. Hensel, U. Lambert, L. Köster: *Mater. Sci. Eng.* **B36**, 33 (1996)
- 5.116 M. W. Qi, S. S. Tan, B. Zhu, P. X. Cai, W. F. Gu, M. Xu, T. S. Shi, D. L. Que, L. B. Li: *J. Appl. Phys.* **69**, 3775 (1991)
- 5.117 A. Gali, J. Miro, P. Deak, C. Ewels, R. Jones: *J. Phys. Condens. Mat.* **8**, 7711 (1996)
- 5.118 W. v. Ammon: Crystal growth of large diameter CZ Si crystals, *Proc. 2nd Int. Symp. on Advanced Science and Technology of Silicon Materials*, Kona, Hawaii 1996, ed. by M. Umeno (145th Committee of the Japan Society for the Promotion of Science, Kona 1996) 233
- 5.119 M. Suhren, D. Gräf, U. Lambert, P. Wagner: Crystal defects in highly boron doped silicon, *Proc. 4th Int. Symp. on High Purity Silicon*, San Antonio 1996, ed. by C. L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H. J. Dawson (The Electrochemical Society, Pennington 1996) 132
- 5.120 T. Sinno, H. Susanto, R. Brown, W. v. Ammon, E. Dornberger: *Appl. Phys. Lett.* **75**, 1544 (1999)
- 5.121 T. Abe, T. Masui, H. Harada, J. Chikawa: In: *VLSI Science and Technology, 1985*, ed. by W. M. Bullis, S. Broyda (The Electrochemical Society, Pennington 1985) p. 543
- 5.122 R. Takeda, T. Minami, H. Saito, Y. Hirano, H. Fujimori, K. Kashima, Y. Matsushita: Influence of LSTD size on the formation of denuded zone in hydrogen-annealed CZ silicon wafers, *Proc. 6th Int. Symp. on High Purity Silicon*, Phoenix 2000, ed. by C. L. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, H. J. Dawson (The Electrochemical Society, Pennington 2000) 331
- 5.123 S. Kishino, M. Kanamori, N. Yoshihizo, M. Tajima, T. Iizuka: *J. Appl. Phys.* **50**, 8240 (1978)
- 5.124 T. Fukuda: *Appl. Phys. Lett.* **65**(11), 1376 (1994)
- 5.125 F. Shimura: *J. Appl. Phys.* **59**, 3251 (1986)
- 5.126 K. Sueoka, M. Akatsuka, K. Nishihara, T. Yamamoto, S. Kobayashi: *Mater. Sci. Forum* **196–201**, 1737 (1995)
- 5.127 J. Vanhellefont, C. Claeys: *J. Appl. Phys.* **62**(9), 3960 (1987)

Diffusion in S

6. Diffusion in Semiconductors

Atomic diffusion in semiconductors refers to the migration of atoms, including host, dopant and impurities. Diffusion occurs in all thermodynamic phases, but the solid phase is the most important in semiconductors. There are two types of semiconductor solid phase: amorphous (including organic) and crystalline. In this chapter we consider crystalline semiconductors and describe the processes by which atoms and defects move between lattice sites. The emphasis is on describing the various conditions under which diffusion can occur, as well as the atomic mechanisms that are involved, rather than on tabulating data. For brevity's sake, we also focus on the general features found in the principal semiconductors from Groups IV, III–V and II–VI; IV–VI and oxide semiconductors are excluded from consideration. It is not surprising that most of the data available in this field relate to the semiconductors that are technologically important – they are used to fabricate electronic and optoelectronic devices. One unavoidable consequence of this technological need is that diffusion data tend to be acquired in a piecemeal fashion.

6.1	Basic Concepts	122
6.2	Diffusion Mechanisms	122
6.2.1	Vacancy and Interstitial Diffusion Mechanisms	122
6.2.2	The Interstitial–Substitutional Mechanism: Dissociative and Kick-Out Mechanisms	122
6.2.3	The Percolation Mechanism	123
6.3	Diffusion Regimes	123
6.3.1	Chemical Equilibrium: Self- and Isoconcentration Diffusion ...	123
6.3.2	Chemical Diffusion (or Diffusion Under Nonequilibrium Conditions).....	123
6.3.3	Recombination-Enhanced Diffusion	125
6.3.4	Surface Effects	125
6.3.5	Short Circuit Paths	125
6.4	Internal Electric Fields	126
6.5	Measurement of Diffusion Coefficients ...	126
6.5.1	Anneal Conditions	126
6.5.2	Diffusion Sources	126
6.5.3	Profiling Techniques	126
6.5.4	Calculating the Diffusion Coefficient.....	127
6.6	Hydrogen in Semiconductors	127
6.7	Diffusion in Group IV Semiconductors	128
6.7.1	Germanium	128
6.7.2	Silicon	128
6.7.3	$\text{Si}_{1-x}\text{Ge}_x$ Alloys	129
6.7.4	Silicon Carbide	129
6.8	Diffusion in III–V Compounds	130
6.8.1	Self-Diffusion	130
6.8.2	Dopant Diffusion	130
6.8.3	Compositional Interdiffusion	131
6.9	Diffusion in II–VI Compounds	131
6.9.1	Self-Diffusion	132
6.9.2	Chemical Self-Diffusion.....	132
6.9.3	Dopant Diffusion	132
6.9.4	Compositional Interdiffusion	132
6.10	Conclusions	133
6.11	General Reading and References	133
	References	133

Diffusion describes the movement of atoms in space, primarily due to thermal motion, and it occurs in all forms of matter. This chapter focuses on atom diffusion in crystalline semiconductors, where diffusing atoms migrate from one lattice site to another in the semiconductor crystal. The diffusion of atoms and defects is at the heart of material processing, whether at the growth or post-growth stage, and control over diffusion

is the basis of process simulation and defect engineering. Such control calls for an understanding of the diffusion processes involved in a given situation. The needs of device technology have provided the main impetus for investigations into the diffusion of atoms in semiconductors. As the physical dimensions of devices have shrunk, the barriers to understanding diffusion mechanisms and processes in complex structures have greatly multiplied.

6.1 Basic Concepts

Consider a particle in a three-dimensional isotropic lattice which migrates by making jumps from one lattice site to a nearest neighbor site. If the distance between nearest neighbor sites is a and the particle makes n jumps in time t , then, assuming each jump is random (so the directions of successive jumps are independent of each other), the mean square displacement $\langle R^2 \rangle$ is equal to na^2 [6.1]. Fick's first law defines the associated diffusivity D to be $na^2/6t = \langle R^2 \rangle/6t = va^2/6$, where $v = n/t$ is the average jump rate of the particle. Taking the diffusion length as $2\sqrt{Dt}$, it follows that this is also equal to $2\sqrt{\langle R^2 \rangle/6}$. For $D = 10^{-12} \text{ cm}^2/\text{s}$, $t = 10^4 \text{ s}$

and $a = 2 \times 10^{-8} \text{ cm}$, the diffusion length is $2 \mu\text{m}$, $n = 1.5 \times 10^8$ jumps and the total distance na traveled by the particle is 3 cm . However, it turns out that in most diffusion mechanisms successive jumps are correlated, not random. The effect of nonrandom jumps is to decrease the diffusivity of the particle relative to what it would be if the jumps were random. Taking this correlation into account leads to $D = fva^2/6$, where $f (\leq 1)$ is the correlation factor [6.1], v is temperature-dependent and f may or may not be, depending on the particular situation. Overall, the temperature dependence of D is found to obey the Arrhenius relation $D = D_0 \exp(-Q/kT)$.

6.2 Diffusion Mechanisms

Two categories of diffusion mechanism are recognized: defect and nondefect. A simple example of the latter class is the simultaneous jumps of two adjacent atoms in order to exchange sites. There is a general consensus, however, that nondefect mechanisms do not play any significant role in semiconductor diffusion, although recently nondefect contributions have been proposed for self- and dopant diffusions in Si [6.2]. In the elemental semiconductors Si and Ge, vacancies and interstitials are the primary defects. In binary compound semiconductors (such as GaAs and ZnSe) there are two sublattices, the anion and cation, so there are vacancies and interstitials for each sublattice together with antisite defects on each sublattice. Further complexity arises due to the various states of ionization of the defects: the mobility of a defect depends on its charge state. The relative concentrations of the different charge states will be determined by the position of the Fermi level.

6.2.1 Vacancy and Interstitial Diffusion Mechanisms

In the Si lattice, a vacancy V_{Si} can migrate by a nearest neighbor Si atom jumping into the vacancy (in other words, the Si atom and V_{Si} have exchanged sites so that the Si atom has also migrated). Equally, a substitutional dopant atom can migrate by jumping into a V_{Si} at a nearest neighbor site. Similarly, in a binary semiconductor such as GaAs, Ga atoms can migrate over the Ga sublattice via jumps into nearest neighbor Ga vacancies, as can dopant atoms substituting into the Ga sublattice. Anti-site defects can diffuse by jumping into vacancies

in the same sublattice, such as the As anti-site defect in GaAs, As_{Ga} , diffusing by jumps into Ga vacancies. For self-interstitials, such as Si_i or Ga_i , their concentrations are sufficiently small for neighboring interstitial sites to always be empty, which means that the occupancy of nearest neighbor sites is not a factor when determining jump rates. If a self-interstitial, such as Si_i , pushes a Si atom on a normal lattice site into an interstitial site instead of jumping into a neighboring interstitial site, and therefore replaces the displaced Si atom, the process is known as the “interstitialcy mechanism”. This concept extends to a substitutional dopant atom forming a pair with a self-interstitial, which then migrates with the dopant atom, alternating between substitutional and interstitial sites. It is also possible for point defects to form complexes which can diffuse as a single entity. Examples are the Frenkel pair $V_{\text{Si}}\text{Si}_i$, di-vacancies such as $V_{\text{Ga}}V_{\text{Ga}}$ or $V_{\text{Ga}}V_{\text{As}}$, and the defect pair formed between a substitutional dopant atom and an adjacent vacancy.

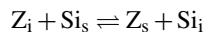
6.2.2 The Interstitial–Substitutional Mechanism: Dissociative and Kick-Out Mechanisms

The interstitial–substitutional diffusion mechanism arises when a dopant species Z occupies both interstitial and substitutional sites, represented by Z_i and Z_s respectively, and diffusion is restricted to jumps of Z_i . In this case, we may ask how the Z_s concentration $[Z_s]$ is linked to the Z_i migration. Consider the diffusion of Z in Si. The dissociative mechanism (also known as the Frank–Turnbull mechanism) is based on the defect

interaction



and application of the law of mass action (LMA) leads to $[Z_i][V_{Si}] \propto [Z_s]$. In the kick-out mechanism, the defect interaction is



and therefore $[Z_i] \propto [Z_s][Si_i]$ ($[Si_s]$ is omitted because it is effectively constant). In order to sustain growth in $[Z_s]$ by either mechanism, it is clearly necessary to have either a supply of V_{Si} or a means of removing Si_i . For simplicity, neutral charge states have been assigned to all of the defects in these two interactions. A detailed treatment of the kick-out mechanism has been given by Frank et al. [6.3]. For in-diffusion of Z, the Frank–Turnbull mechanism consumes vacancies and will therefore tend to reduce the local vacancy concentration, whereas the

local self-interstitial concentration will be enhanced by the kick-out mechanism. Out-diffusion of Z reverses the effects on the native defect concentrations.

6.2.3 The Percolation Mechanism

The percolation mechanism [6.4] was proposed to explain group V dopant diffusion in Si at high dopant concentrations (in excess of $\approx 1\%$). At low concentrations diffusion is via dopant- V_{Si} pairs. As the dopant concentration increases, regions occur in the Si lattice where the proximity of the dopant atoms enhances the mobility and concentration of the V_{Si} . Within this network the diffusivity of dopant- V_{Si} pairs is thereby also enhanced so that the dopant diffusivity increases overall. The percolation network only forms once the dopant concentration exceeds a certain critical value. In principle this mechanism could extend to other highly doped materials.

6.3 Diffusion Regimes

The mobility of a native defect and/or dopant atom reflects the physical and chemical environment under which diffusion is occurring. Two types of environment arise: conditions of chemical equilibrium and those of chemical nonequilibrium. Diffusion in temperature gradients is excluded – only isothermal conditions are considered.

6.3.1 Chemical Equilibrium: Self- and Isoconcentration Diffusion

Chemical equilibrium means that the concentrations of all chemical components, including native defects, are uniform throughout the semiconductor, and where appropriate (such as in a compound material), the solid is in equilibrium with the ambient vapor of the components so that the level of nonstoichiometry is defined. Experimentally this requires diffusion to be carried out in a sealed system. Self-diffusion refers to the diffusion of the host atoms, such as Si atoms diffusing in the Si lattice. Isoconcentration diffusion describes the diffusion of dopant atoms when the same dopant concentration is uniform throughout the sample, such as for As diffusion in Si for a constant As doping level. In either case, diffusion can only be observed if some of the particular diffusing atoms are tagged, such as by using a radioisotope or an isotopically enriched diffu-

sion source. The diffusivity of a tagged or tracer species is related to the concentration of the native defect that provides the diffusion path, and the self-diffusivity (the diffusivity of the tracer) is always significantly smaller than the associated defect diffusivity. Specific relations can be found in Shaw [6.5]. A diffusion flux of the tracer arises from a tracer concentration gradient, which is not to be confused with a chemical concentration gradient.

6.3.2 Chemical Diffusion (or Diffusion Under Nonequilibrium Conditions)

This category contains all of the diffusion phenomena that are of technological interest and importance. In this case, diffusion occurs due to spatial gradients in the concentrations of the chemical components in the material, which are in turn caused by departures from equilibrium: the diffusion processes are attempting to either restore or achieve equilibrium.

Chemical Self-Diffusion

Chemical self-diffusion describes the process whereby a compound semiconductor changes from one level of nonstoichiometry to another through changes in the native defect populations. These changes can arise due to a change in the ambient partial pressure of one

of the components (that of As in the case of GaAs for example), or through a change in temperature under a defined or controlled component partial pressure. Good examples are provided by II–VI semiconductors [6.6] and in particular $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ [6.7], where p- to n-type conversion is used to form p–n junctions by annealing in Hg vapor. Changes in the native defect concentrations can also lead to an increased dopant diffusivity; an increase in the vacancy concentration will enhance any diffusivity based on a vacancy mechanism for instance.

Dopant Diffusion

Suppose we have a dopant diffusing into the semiconductor from a source located at an external surface (such as a surface layer) or in an external phase (such as a gas or vapor). The existence of the dopant concentration gradient can lead to various effects which can influence the dopant flux. For example, if a dopant diffuses via a vacancy mechanism, then at any position in the diffusion region the increase in the dopant concentration requires a supply of vacancies, so that to maintain local defect equilibrium there must also be an associated vacancy flux. If the dopant controls the position of the Fermi level, then the concentrations of ionized native defects will increase or decrease, depending on their charge state, relative to their intrinsic concentrations (the Fermi-level effect). This means that the concentrations of native defects of opposite (the same) polarity to the dopant will be increased (decreased). Increases in the concentrations of ionized native defects due to Zn, Si or Te diffusion into GaAs/GaAlAs superlattice structures explain the disordering of the superlattices [6.8]. Usually a substitutional dopant atom will have a different size to that of the host atom it replaces. This size difference creates a local mechanical strain which in turn can cause changes in the local concentrations of native defects as well as to jump rates and hence dopant diffusivity [6.9]. If the strain is large enough, misfit dislocations will be generated [6.10], otherwise there will be a strain energy gradient matching the dopant concentration gradient which can enhance or retard the dopant flux [6.11]. More recently a new scenario has emerged: dopant diffusion in strained epilayers. Whether diffusion is enhanced or retarded depends on several poorly understood parameters [6.12]. However, significant effects are found, such as the decrease in the B diffusivity in strained $\text{Si}_{1-x}\text{Ge}_x$ epilayers by a factor of ≈ 10 as the strain increased from zero to 0.64 [6.13].

Compositional Interdiffusion (CID)

Compositional interdiffusion describes diffusion across the interface separating two materials of different chemical composition. Chemical composition here refers to major components; dopants and deviations from stoichiometry are excluded. CID can be exploited when making graded bandgap structures and during material preparation, such as in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, where an alternating sequence of HgTe and CdTe epilayers of appropriate thicknesses are first grown and then interdiffused. CID can also pose problems in the fabrication of multiple quantum wells and superlattice structures when sharp boundaries are necessary. In particular, donor or acceptor dopant diffusion into GaAs-based superlattices can cause essentially complete intermixing on the cation sublattice [6.8]. This phenomenon is also known as diffusion-induced disorder. B or As doping also results in rapid intermixing at a Si/Ge interface [6.14].

Transient Enhanced Diffusion (TED)

Ion implantation is often the preferred way to achieve a doped layer. The implantation process does however create a significant amount of lattice damage so that a subsequent anneal stage is needed in order to achieve full electrical activity of the implanted dopant and recovery of the lattice damage. During the implantation process, the implant ions create collision cascades of vacancies and self-interstitials (an excess of native defects). The post-implant anneal serves to remove or reduce this excess. In addition to vacancy/interstitial recombination, the excess native defects can interact to form clusters (which also may contain implant ions as well as residual impurities) and extended defects, such as dislocation loops. At the start of the post-implant anneal, the local concentrations of vacancies and self-interstitials in the implant region can greatly exceed equilibrium values and therefore enhance the implant ion diffusivity in the implant region. As annealing proceeds the excess concentrations will diminish and will be reflected in a diminishing dopant diffusivity until values appropriate to local equilibrium are reached. This temporary enhancement in the dopant diffusivity is known as TED. The topic is a complex one to analyze quantitatively and detailed consideration of the issues involved in the case of B implants in Si can be found in the review by Jain et al. [6.15]. In the fabrication of shallow p–n junctions using ion implants and rapid thermal annealing (RTA), TED can determine the lower limit to junction depth. TED of B in Si can be reduced with coimplants of Si prior to RTA [6.16]. TED of Be and Si

in GaAs has also been discussed [6.17]. Some workers use the term **TED** to describe the enhanced diffusivity of a dopant, incorporated during growth, which occurs when the structure is annealed at a higher temperature than the growth temperature, such that there is an initial supersaturation of the relevant native defects at the anneal temperature.

Segregation, Gettering, Precipitation and Clustering

A variety of important scenarios arise, involving many of the above regimes, during growth and/or thermal processing stages of materials and structures. The segregation of acceptor dopants in InP [6.18] and in III–V heterostructures [6.19, 20] has been observed and modeled. The segregation (or accumulation) due to diffusion of the dopant is in effect a partitioning process to preferred (higher solubility) regions within the layer structure. Gettering describes the segregation, or clean-up, of a fast-diffusing impurity from the active regions of a device structure. Such impurities are typically Group IB and transition metals and are incorporated either during growth or during subsequent processing. Gettering sites in Si are provided by O precipitates, self-ion implant damage layers and nanocavities [6.21]. In contrast, Group IB impurities are gettered in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ by regions of high cation vacancy concentration [6.22].

Precipitation occurs when a species – whether native defect, dopant or impurity – becomes supersaturated, and in order to achieve its equilibrium concentration the species excess is removed by the formation of precipitates within the host lattice. A self-interstitial or vacancy excess can be removed through the nucleation of dislocation loops, stacking faults or voids, which then provide sites for the precipitation of the remaining excess. In the case of a dopant, impurity or nonstoichiometric excess, nucleation of a precipitate can be spontaneous (homogeneous) or heterogeneous. The latter occurs at the site of an impurity atom (for example, C atoms in Si serve as nucleation centers for the precipitation of O) or at dislocations, giving rise to the term “decoration”. Growth of any precipitate proceeds via diffusion of the particular species from solution in the matrix to the precipitate and is generally diffusion-limited. Invariably local stress fields will be present which influence the diffusion and, if present initially, they may also play a role in the nucleation stage. The precipitation of O impurities in Si presents a unique case study because of the high [O], its technological importance and its complexity [6.23, 24]. The rather simpler case of B precipitation in Si has

been described by Solmi et al. [6.25]. A cluster (or agglomerate) refers to a configuration of at least a few dopant atoms (with or without associated native point defects) or host species. Solmi and Nobili [6.26] have identified $(2\text{As}_{\text{Si}} - \text{V}_{\text{Si}})^0$ and $(4\text{As}_{\text{Si}} - \text{V}_{\text{Si}})^+$ clusters in heavily As-doped Si. Heavy C doping [6.27] and B implants [6.28] in Si give rise to self-interstitial clusters with C and B respectively. In Si, according to Ortiz et al. [6.29], if the number of self-interstitials in a cluster exceeds ≈ 10 there is a transition to a {113} defect.

6.3.3 Recombination-Enhanced Diffusion

The local energy released in the nonradiative recombination of excess free carriers can help a diffusing species to surmount the energy barrier separating it from an adjacent lattice site – in other words, the energy barrier facing a jump is effectively reduced. This situation is important in the degradation of performance of device structures which utilize high excess minority carrier concentrations, such as light-emitting and laser diodes.

6.3.4 Surface Effects

The concentrations of native point defects within the bulk can be altered by surface processes. In the case of Si it is well known that during surface oxidation or nitridation there is injection of Si interstitials or of vacancies respectively. All diffusants can therefore be affected during the duration of the process. Ion beam milling causes the injection of Hg interstitials into $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ in sufficient quantities to effect p- to n-type conversion.

6.3.5 Short Circuit Paths

The existence of dislocations and subgrain boundaries in single-crystal materials generally provides high diffusivity routes for all atomic species relative to the surrounding matrix. Care is always needed when evaluating experimental data to ensure that bulk diffusion is not being masked by short circuit paths [6.30]. In the case of polycrystalline Si, the grain boundaries may provide high diffusivity routes, as in the cases of As and B [6.31], or retard diffusion, as for Au [6.32]. The situation is a complex one, as grain growth also occurs during any anneal. Kaur et al. [6.33] have provided a comprehensive account of short circuit path diffusion.

6.4 Internal Electric Fields

When the dopant concentration is large enough to make the diffusion zone electrically extrinsic, free carriers from the dopant, due to their much higher mobility, will diffuse ahead of the parent dopant atoms. This separation creates a local electric field whose direction is such as to pull the dopant atoms after the free carriers (and also to pull the free carriers back). Provided that the diffusion length $> \approx$ six Debye screening lengths (typically $\approx 10^2$ nm), the diffusion zone can be regarded as electrically neutral (the space charge density is negligible) [6.34]. In this situation the local electric field E is given by $-(kT/en)(\partial n/\partial x)$ for an ionized donor dopant diffusing parallel to the x -axis, and nondegenerate conditions apply: k , T , e and n are Boltzmann's constant, the absolute temperature, the electronic charge and the free electron concentration respectively. E exerts a force on each ionized donor (D^+) parallel to the x -axis, creating a local donor flux $-(D(D^+)/n)(\partial n/\partial x)$ due to drift in the electric

field: $D(D^+)$ is the donor diffusivity [6.35]. This drift flux adds to the diffusion flux, $-D(D^+)\partial[D^+]/\partial x$, to give the total donor flux at any position in the diffusion region, so that the donor flux in this case is increased due to E . E will also cause drift of any other charged species.

Internal electric fields can arise in other circumstances such as in depletion layers where E must be calculated from Poisson's equation, in graded bandgap structures [6.11, 36], and at the interfaces of heterostructures. Cubic II–VI and III–V strained layer heterostructures grown on the $\{111\}$ direction are piezoelectric and typical strains from lattice mismatch of $\approx 1\%$ can give $E \approx 10^5$ V/cm in the absence of free carrier screening [6.37]. In wurtzite heterostructures based on the Ga, In nitrides, even higher fields are found ($E \approx 10^6$ V/cm) due to piezoelectric and spontaneous polarization [6.38]. These fields can be important in CID and chemical self-diffusion.

6.5 Measurement of Diffusion Coefficients

6.5.1 Anneal Conditions

Accurate control of sample temperature and ambient are essential if controlled and reproducible results are to be obtained in a diffusion anneal. Depending on the time spent at the anneal temperature, the warm-up and cool-down times may also be important. An appropriate choice of ambient is needed to preserve the sample surface (to avoid evaporation, surface melting or alloying with the dopant source for example). For compound semiconductors it is necessary to define the level of non-stoichiometry by controlling the ambient partial pressure of one of the components, such as As for GaAs or Hg for $Hg_{1-x}Cd_xTe$. If the dopant is in an external phase, knowledge of the phase diagram of all of the components is required [6.1, 35]. Control over partial pressure is best achieved in a sealed system, typically a fused silica ampoule. Annealing in a vertical or horizontal resistance-heated furnace requires a minimum anneal time of 30 to 60 min in order to avoid uncertainties due to warm-up and cool-down. The drive to shallow dopant profiles has been facilitated through rapid thermal annealing (RTA) techniques. These are based on radiant heating of the sample, and linear heating rates of $100\text{--}400^\circ\text{C/s}$ with cooling rates of up to 150°C/s are available. RTA however precludes the use of a sealed

system and, in this case, a popular means of preventing surface deterioration is to seal the sample with an inert, impervious capping layer, made of silicon nitride for example.

6.5.2 Diffusion Sources

Consideration is limited to planar samples with diffusion normal to a principal face. This is a common situation and diffusion of a dopant or tracer species can take place from: (i) a surrounding vapor or gas phase; (ii) a surface layer, which may be evaporated, chemically deposited (CVD) or a spun-on silicate glass, all incorporating the diffusant; (iii) epilayers containing the diffusant, which may provide the external surface or be buried within the epitaxial structure; (iv) ion implants of a dopant either directly into the sample surface or into a thin surface layer so as to avoid lattice damage. It is obviously desirable that negligible diffusion occurs prior to reaching the anneal temperature when the diffusant is incorporated into an epilayer. In self-diffusion experiments the tracer can be a radiotracer or an isotopically enriched species. A key requirement for either form of tracer is availability, and a radiotracer must have a half-life that is long enough for the experiments to be carried out.

6.5.3 Profiling Techniques

Determining the spatial distribution of a diffusant for various anneal times is fundamental to obtaining its diffusion coefficient or diffusivity. Most methods are destructive, as they generally require a bevel section through the diffusion zone or the sequential removal of layers. The two broad profiling categories are electrical and species-specific. Electrical methods are primarily the p–n junction method, spreading resistance and capacitance–voltage profiling. Limitations of the electrical methods are: (i) assumptions are needed to link the electrical data to the diffusant (for example, that the diffusant is the only electrically active center and that it is fully ionized); (ii) the assumption that the anneal temperature defect situation is “frozen-in” during cool-down. Electrical methods are the most direct means of measuring chemical self-diffusivities and can readily detect changes in host concentrations of < 1 part in 10^4 . Species-specific (chemical element or isotope) profiling means that the chemical concentration of the diffusant is determined regardless of its location(s) in the lattice and of its electrical state. Profiling of the diffusant using a radiotracer has been widely used [6.39], but in the past decade or so secondary ion mass spectrometry (SIMS) has become what is essentially the standard procedure for diffusant profiling. This is because SIMS can measure diffusant concentrations within the range 10^{16} to 10^{22} cm $^{-3}$ with spatial resolutions at best of several nanometers per decade (of concentration). Primary factors determining the resolution are progressive roughening of the eroded surface and “knock-on” effects due to the probing ion beam displacing the diffusant to greater depths. A further problem may arise when the atomic mass of a dopant is close to that of the host species.

Nondestructive profiling techniques applicable to CID in quantum well and superlattice structures utilize either high-resolution X-ray diffraction

(HRXRD) [6.40] or photoluminescence (PL). The detail in the X-ray diffraction patterns reflects the CID profiles at the interfaces and can also reveal the presence of strain in the structures. The use of PL requires the presence of optically active centers in the quantum well. CID changes the shape and depth of the quantum well, which in turn changes the photon energies in the luminescence spectra. HRXRD and PL can also be combined. A particular advantage of these techniques is that they allow successive anneals to be performed on the same sample.

6.5.4 Calculating the Diffusion Coefficient

Once a planar concentration profile has been obtained, the first step is to see if the profile can be fitted to a solution of Fick’s second law. The simplest solution occurs for a diffusivity D independent of the diffusant concentration (c), for a constant surface concentration c_0 and a diffusion length \ll the layer or sample thickness. The solution is $c = c_0 \operatorname{erfc}[x/2\sqrt{(Dt)}]$ [6.1]. If the profile is not erfc, it may be because D varies with c , and D (as a function of c) can be obtained by a Boltzmann–Matano analysis [6.35]. It is important to recognize that the erfc or Boltzmann–Matano solutions are only valid provided c_0 does not change with time and that c/c_0 versus x/\sqrt{t} for profiles at various t reduce to a single profile. More complex situations and profiles require numerical integration of the appropriate diffusion equation(s) and matching to the experimental c versus x profile; in other words a suitable model with adjustable parameters is used to simulate the observed profiles. The interpretation of HRXRD and/or PL data provides a good example of a simulation scene in which an assumed D , either c -dependent or c -independent, is used to calculate the resulting CID profile and its effect on the X-ray patterns and/or PL spectra. Whereas SIMS can observe diffusivities as low as $\approx 10^{-19}$ cm 2 /s, the HRXRD limit is $\approx 10^{-23}$ cm 2 /s.

6.6 Hydrogen in Semiconductors

Hydrogen is a ubiquitous element in semiconductor materials and can be incorporated either by deliberate doping or inadvertently, at significant concentrations, during growth and/or in subsequent surface treatments where organic solvents, acid or plasma etching are used. H is known to passivate electrically active centers by forming complexes with dopants and native defects as well as by bonding to the dangling bonds at extended

defects. Such interactions may well affect the diffusivities of the dopant and native defect. This expectation is realized in the case of O in Si, where the presence of H can enhance O diffusivity by two to three orders of magnitude [6.41]. Ab initio calculations show that, at least in the Group IV and III–V semiconductors, H is incorporated interstitially in the three charge states, H^+ , H^0 and H^- , with the Fermi level controlling

the relative concentrations. In addition to interactions with dopant atoms and native defects, H_2 molecules also form. Mathiot [6.42] has modeled H diffusion in terms of simultaneous diffusion by the three interstitial charge

states with the formation of immobile neutral complexes. In polycrystalline Si, the grain boundaries retard H diffusion, so H diffuses faster in the surrounding lattice than in the grain boundary.

6.7 Diffusion in Group IV Semiconductors

Diffusants divide into one of five categories: self-, other Group IVs, slow diffusers (typically dopants from Groups III and V), intermediate diffusers and fast diffusers. The materials of interest are Ge, Si, Si/Ge alloys and SiC. A particular feature is that self-diffusion is always slower than the diffusion of other diffusants. With the exception of SiC, which has the zinc blende structure, as well as numerous polytypes (the simplest of which is the wurtzite, $2H-SiC$, form), the other members of this group have the diamond lattice structure.

6.7.1 Germanium

The evidence to date identifies the dominant native defect in Ge as the singly ionized vacancy acceptor, V_{Ge}^- [6.43], which can account for the features found in self-diffusion and in the diffusivities of dopants from groups III and V. The self-diffusivity, relative to the electrically intrinsic value, is increased in n-type Ge and decreased in p-type as expected from the dependence of $[V_{Ge}^-]$ on the Fermi level. In intrinsic Ge the best parameters for the self-diffusivity are $D_0 = 13.6 \text{ cm}^2/\text{s}$ and $Q = 3.09 \text{ eV}$, from Werner et al. [6.44], because of the wide temperature range covered (535–904 °C). The diffusivities of donor dopants (P, As, Sb) are very similar in magnitude, as are those for acceptor dopants (Al, In, Ga). The acceptor group diffusivities, however, are very close to the intrinsic self-diffusivity, whereas those for the donor group are 10^2 to 10^3 times larger. Li is a fast (interstitial) diffuser with a diffusivity exceeding the donor group diffusivities by factors of 10^7 to 10^5 between 600 and 900 °C, whereas Cu [6.45] and Au [6.46] are intermediate (dissociative) diffusers.

6.7.2 Silicon

Si stands alone due to the intensive investigations that have been lavished on it over the past 50 years. In the early days diffusion data yielded many perplexing features. Today the broad aspects are understood along with considerable detail, depending on the topic. Diffusion in Si covers many more topics than arise in any other semi-

conductor and it is still a very active area of R & D. It is now recognised that, apart from foreign purely interstitial species, self-interstitials, Si_i , and vacancies, V_{Si} , are involved in all diffusion phenomena. So far the best self-diffusion parameters obtained for intrinsic Si are $D_0 = 530 \text{ cm}^2/\text{s}$ and $Q = 4.75 \text{ eV}$ in the temperature range 855–1388 °C [6.47]. Two distinct facets of self- and dopant diffusion in Si are: (a) the diffusivity has two or three components, each with differing defect charge states; (b) the diffusivity reflects contributions from both Si_i and V_{Si} [6.2, 43, 48, 49]. Thus the Si self-diffusivity is determined by Si_i and V_{Si} mechanisms and by three separate defect charge states: neutral (0), positive (+) and negative (−). Identifying which charge state goes with which defect remains a problem. For the common dopants (B, P, As and Sb), B and P diffuse primarily via the Si_i defect, As diffuses via both Si_i and V_{Si} defects, whereas Sb diffuses primarily via V_{Si} . Two defect charge states are involved for B (0, 1+), As (0, 1−) and Sb (0, 1−), and three for P (0, 1−, 2−). The situation for Al [6.50], Ga [6.9] and In [6.51] has Si_i dominant for Al and In diffusion whereas both Si_i and V_{Si} are involved for Ga. The associated charge states are Al (0, 1+), Ga (0, 1+) and In (0). The diffusivities of the Group V donor dopants (P, As, Sb) lie close to each other and are up to a factor of ≈ 10 greater than the self-diffusivity. The acceptor dopants (B, Al, Ga, In) also form a group with diffusivities that are up to a factor of $\approx 10^2$ greater than the donor dopants. A recently observed interesting feature is that the diffusivities of B and P in intrinsic material depend on the length of the anneal time, showing an initial change until reaching a final value [6.52]. This time effect is attributed to the time needed for equilibration of the V_{Si} and Si_i concentrations at the anneal temperature.

The data presented by Tan and Gösele [6.43] show that Au, Pt and Zn are intermediate (kick-out) diffusers and that H, Li, Cu, Ni and Fe are fast interstitial diffusers. Recent evidence shows that Ir diffusion occurs via both kick-out and dissociative mechanisms [6.53]. To provide some perspective: at 1000 °C the diffusivity of H is $\approx 10^{-4} \text{ cm}^2/\text{s}$ compared to a self-diffusivity

of $8 \times 10^{-17} \text{ cm}^2/\text{s}$. C and O are important impurities because, though electrically neutral, they occur in high concentrations and can affect the electrical properties. Although O occupies interstitial sites and diffuses interstitially it should be classed as an intermediate diffuser because a diffusion jump entails the breaking of two Si–O bonds. C has a diffusivity that is a little larger than those of Group III dopants: its mechanism is unresolved between the “kick-out” mode or a diffusing complex comprising a Si_i and a substitutional C.

6.7.3 $\text{Si}_{1-x}\text{Ge}_x$ Alloys

Si and Ge form a continuous range of alloys in which there is a random distribution of either element as well as a continuous variation of bandgaps. The alloys have attracted considerable interest from a device perspective and are usually prepared as epilayers on Si substrates so that the epilayer will generally be in a strained state. Diffusivity data are sparse and, in the case of dopants, limited to B, P and Sb. One might expect that the diffusivity $D(Z)$ of dopant Z would increase continuously as x goes from 0 to 1 at any given temperature below the melting point of Ge. However, in the case of B, $D(\text{B})$ hardly varies for $x \lesssim 0.4$; even so, $D(\text{B})$ increases by a factor $\approx 10^3$ from $\approx 10^{-15} \text{ cm}^2/\text{s}$ in traversing the composition range at 900°C [6.54, 55]. $D(\text{P})$ increases by a factor of ≈ 4 for x values between 0 and 0.24, only to show a decrease at $x = 0.40$ [6.55]. Limited data suggest that $D(\text{Sb})$ rises continuously across the composition range, increasing by a factor $\approx 10^6$ at 900°C [6.56]. Surface oxidation enhances $D(\text{B})$ and $D(\text{P})$, indicating that the diffusivities are dominated by a self-interstitial mechanism, whereas $D(\text{Sb})$ is reduced by surface oxidation, pointing to a vacancy mechanism. Compressive strain retards $D(\text{B})$ whereas tensile strain gives a marginal enhancement [6.55]. Compressive strain enhances $D(\text{P})$ and $D(\text{Sb})$ [6.57]. Overall, some disagreement exists between different workers about the behavior of $D(Z)$, which may well stem from difficulties with characterizing the experimental conditions. Compositional interdiffusion has been characterized at the interface between Si and layers with $x < 0.2$ [6.58].

6.7.4 Silicon Carbide

Its large bandgap, high melting point and high dielectric breakdown strength make SiC a suitable material for devices intended for operation at high temperatures and high powers. It also has potential optoelectronic appli-

cations. Characterizing the material is complicated, as SiC occurs in a range of polytypes (different stacking sequences of close packed layers). Common polytypes are the cubic zinc blende phase 3C-SiC and the hexagonal phases 2H-SiC (wurtzite), 4H-SiC and 6H-SiC . This combination of high melting point, polytypism and variations in stoichiometry makes it difficult to measure diffusivities. Typical diffusion anneal temperatures for acceptor (B, Al, Ga) and donor (N, P) dopants are in the range $1800\text{--}2100^\circ\text{C}$. Ab initio calculations for single vacancies and anti-sites in 4H-SiC [6.59] found the Si_C and C_Si anti-sites to be both neutral and therefore generally inactive (electrically and optically). The C vacancy is amphoteric with charge states ranging from $2+$ to $2-$. The Si vacancy is also amphoteric with charge states ranging from $1+$ to $3-$. Similar calculations for self-interstitials in 3C-SiC [6.60] predict divalent donor behavior for both Si and C interstitials. Bockstedte et al. [6.61] have calculated, using ab initio methods, the activation energies Q for self-diffusion in 3C-SiC by vacancies and self-interstitials. Generally Q is smaller for self-interstitials but the defect charge state is also an important factor. The Si vacancy is predicted to be metastable, readily transforming to the stable complex $\text{V}_\text{C}-\text{C}_\text{Si}$; the complex $\text{V}_\text{Si}-\text{Si}_\text{C}$ is unstable, reverting to V_C .

The Si and C self-diffusivities, $D(\text{Si})$ and $D(\text{C})$, respectively, were measured between 1850 and 2300°C by Hong et al. ([6.62] and references therein) in both 3C-SiC and 6H-SiC . The ratio $D(\text{C})/D(\text{Si})$ was ≈ 650 in 3C-SiC and ≈ 130 in 6H-SiC . N doping increased $D(\text{Si})$ and reduced (marginally) $D(\text{C})$. This behavior suggests that native acceptors are important for Si self-diffusion and that native donors are only marginally involved in determining $D(\text{C})$. Of particular interest is that, between the two polytypes, the self-diffusivities in 6H-SiC exceeded those in 3C-SiC by less than a factor of ≈ 3 . This suggests that diffusivities are insensitive to the particular polytype. More recent measurements of $D(\text{C})$, between 2100 and 2350°C , in 4H-SiC found diffusivities that were $\approx 10^5$ times smaller than the earlier results for 3C-SiC and 6H-SiC , mainly because of differences in D_0 [6.63]. There is currently no explanation for these huge differences and the question of the reliability of self-diffusivity data must be considered.

Earlier work by Vodakov et al. [6.64] found that the diffusivity of B in six different polytypes of SiC, excluding 3C-SiC , varied by $\leq 30\%$, not only for diffusion along the c -axis but also perpendicular to it. The diffusivities of some common dopants have been sum-

marized by *Vodakov* and *Mokhov* [6.65]. B diffusion mechanisms in 4H and 6H-SiC have been discussed by *Usov et al.* [6.66]. A recent finding is that an SiO₂ layer on the surface of 6H-SiC greatly enhances B diffusion [6.67], yielding a diffusivity of $\approx 6 \times 10^{-16} \text{ cm}^2/\text{s}$

at 900 °C. This compares to a temperature of $\approx 1400 \text{ °C}$ (extrapolated) for the same diffusivity without an SiO₂ layer. Electric fields of $\approx 10^6 \text{ V/cm}$ have been found in 4H/3C/4H-SiC quantum wells due to spontaneous polarization in the 4H-SiC matrix [6.68].

6.8 Diffusion in III–V Compounds

The III–V binary compounds are formed between the cations B, Al, Ga, In and the anions N, P, As and Sb. Mutual solubility gives rise to the ternaries, such as Al_{1-x}Ga_xAs, and to the quaternaries, such as In_{1-x}Ga_xAs_{1-y}P_y. The B compounds offer little more than academic interest, whereas the rest of the III–V family are important materials in both electronic and optoelectronic devices. The nitrides all have the wurtzite structure, with the remaining compounds possessing the zinc blende structure. In view of the wide range of binaries, ternaries, and so on, it is not surprising that diffusivity measurements have focused mainly on those compounds relevant to devices: essentially GaAs and GaAs-based materials. An important characteristic of these compounds is the high vapor pressures of the anion components; it is the variations in these components that lead to significant changes in levels of nonstoichiometry. This means that a proper characterization, at a given temperature, of any diffusivity must specify the doping level and the ambient anion vapor pressure during the anneal: the latter determines native defect concentrations in intrinsic samples, and both factors have equal importance in controlling the concentrations under extrinsic conditions. On both the anion and the cation sublattices, the possible native point defects are the vacancy, the self-interstitial and the anti-site and all can occur in one or more charge states.

and Ga_i²⁺ could dominate at high doping levels. Between 800 and 1200 °C the Arrhenius parameters for $D(\text{Ga})$ are $D_0 = 0.64 \text{ cm}^2/\text{s}$ and $Q = 3.71 \text{ eV}$ in intrinsic GaAs under a partial As₄ vapor pressure of $\approx 1 \text{ atm}$. The situation for As self-diffusion is less clear, but the evidence points to the dominance (in the diffusion process) of the neutral As interstitial over the As vacancy (the supposedly dominant native defect, the As anti-site, is not involved). Data have been obtained for both Ga and Sb self-diffusion in intrinsic GaSb under Ga- and Sb-rich conditions. There is a conflict between the results obtained with bulk material and those from isotope heterostructures (see [6.70] and references therein). *Shaw* [6.70] concluded that the defects involved in Ga self-diffusion were the Frenkel pair Ga_iV_{Ga} and V_{Ga} even though the Ga anti-site Ga_{Sb} appears to be the dominant native defect. Two parallel mechanisms were also identified for Sb self-diffusion, namely one due to the defect pair Sb_iV_{Ga} and the second due to either to the mixed vacancy pair V_{Ga}V_{Sb} or to the triple defect V_{Ga}Ga_{Sb}V_{Ga}. Reliable results for $D(\text{Ga})$ in intrinsic GaP under a partial vapor pressure (P_4) of $\approx 1 \text{ atm}$ are also available [6.71]: between 1000 and 1190 °C the Arrhenius parameters for $D(\text{Ga})$ are $D_0 = 2.0 \text{ cm}^2/\text{s}$ and $Q = 4.5 \text{ eV}$. Data on the effects of doping and changing partial pressure are lacking.

6.8.1 Self-Diffusion

Self-diffusivity data are limited to the Ga and In compounds [6.35, 69], and even here systematic measurements are restricted to GaAs [6.43, 69] and GaSb ([6.70] and references therein). For GaAs, early evidence (based largely on CID in AlGaAs structures) concluded that the Ga self-diffusivity $D(\text{Ga})$ was determined by the triply ionized Ga vacancy V_{Ga}^{3-} and doubly ionized Ga interstitial Ga_i^{2+} . More recent and direct measurements of $D(\text{Ga})$ in Ga isotope heterostructures identified the three vacancy charge states V_{Ga}^{2-} , V_{Ga}^{1-} and V_{Ga}^0 as being responsible for $D(\text{Ga})$ in intrinsic and lightly doped GaAs; the possibility remains that V_{Ga}^{3-}

6.8.2 Dopant Diffusion

Most of the data on dopant diffusion in the III–Vs refer to GaAs [6.35], notably for Be [6.72], Cd [6.69], C, Si, S, Zn and Cr [6.43]. The singly ionized acceptors Be, Zn and Cd (which occupy Ga sites) and the singly ionized donors C and S (which occupy As sites) all diffuse via the kick-out mechanism. The native interstitials involved are Ga_i²⁺ and As_i⁰, apart from Be where the data are best accounted for in terms of the singly ionized interstitial Ga_i¹⁺. Si is an amphoteric dopant and at low concentrations it predominantly occupies Ga sites as a singly ionized donor Si_{Ga}¹⁺. At high concentrations compensation starts to occur due to increasing occupancy as a singly ionized acceptor on As sites. At low

concentrations $\text{Si}_{\text{Ga}}^{1+}$ diffusion is attributed to a vacancy mechanism ($\text{V}_{\text{Ga}}^{3-}$). Cr sits on Ga sites and is a deep-level acceptor dopant important in the growth of high resistivity GaAs. Depending on circumstances, it can diffuse by either the kick-out or the Frank–Turnbull mechanism. The creation of extended defects in the diffusion zone by Zn in-diffusion in GaAs is a well-established feature. The same feature has also been found by *Pöpping* et al. [6.73] for Zn in-diffusion in GaP. They further concluded that Zn diffuses via the kick-out process in GaP through the involvement of either Ga_i^{1+} or Ga_i^{2+} .

6.8.3 Compositional Interdiffusion

The III–V binaries, ternaries and quaternaries are the bases for the fabrication of numerous quantum well and superlattice structures. CID is clearly an issue in the integrity of such structures. The general situation in which the cation and anion sublattices in each layer can contain up to four different components with concentrations ranging from 0 to 100% presents an impossibly complex problem for characterizing diffusion behavior with any rigour. The role of strain in the layers must also be considered a parameter. As a consequence, CID studies have been limited to simpler structures, primarily GaAs–AlAs and GaAs–AlGaAs with interpretations in terms of known diffusion features in GaAs [6.43]. Doping is an important ingredient of these multilayer structures and it was soon discovered that the acceptors Be, C, Mg, Zn and the donors Si, Sn, S, Se and Te could all cause complete disorder of the structure through enhancement of the CID process on either or both sublattices [6.43, 74]. An interesting exception,

however, is found in GaAs–GaAsSb, where either Si or Be reduce CID. Two generally accepted reasons for these dopant effects are: (i) the Fermi-level effect in which the dopant (acceptor/donor) concentration is high enough to make the semiconductor extrinsic so that the concentrations of native (donor/acceptor) defects are increased; (ii) if the dopant diffuses by the kick-out mechanism then in-diffusion will generate a local excess of the native self-interstitial. Clearly (i) operates for dopants incorporated during growth or by subsequent in-diffusion, whereas (ii) is restricted to in-diffusion. Either way the increase in the local native defect concentration(s) leads to a direct enhancement of CID. In the case of GaAs–GaAsSb, cited above, Si will also decrease the concentrations of native donors such as native anion vacancies, which would have a direct impact on and reduction of CID on the anion sublattice. On the other hand, Be should increase native donor concentrations and therefore give enhanced CID of the anions, contrary to observation. Overall, the general features of the dopant-induced disordering process seem to be understood but problems still remain. *Harrison* [6.74] has commented on the approximations commonly made when extracting quantitative information from CID data. The demands of III–V device technology present increasing complexity when attempting to understand the physical processes involved, so that recourse to empirical recipes is sometimes needed. This is illustrated by structures comprising GaInNAs quantum wells with GaAs barriers, all enclosed within AlAs outer layers, whose optoelectronic properties can be improved by the judicious choice of time/temperature anneals [6.75].

6.9 Diffusion in II–VI Compounds

Interest in II–VI materials pre-dates that in the III–Vs because of their luminescence properties in the visible spectrum, which, based on powder technology, resulted in the application of the bigger bandgap materials (such as ZnS) as phosphors in luminescent screens. The development of crystal growth techniques extended interest in the optoelectronic properties of the wider family of II–VI binary compounds formed between the group II cations Zn, Cd and Hg and the group VI anions S, Se and Te. As with the III–Vs, ternary and quaternary compounds are readily formed. The ternary range of compositions $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ has proved to be the most important family member because of their unique properties and consequent extensive exploitation in infra-red systems. ZnS,

CdS and CdSe crystallize in the wurtzite structure, whereas the remaining binaries have the zinc blende structure. The native point defects that can occur are similar to the III–Vs; namely, vacancies, self-interstitials and anti-sites for the cation and anion sublattices. Recent interest has expanded to include the cations Be, Mg and Mn, usually in ternary or quaternary systems. A distinctive feature of atomic diffusion in the II–VI compounds is the much higher diffusivities relative to those in the Group IV and III–V semiconductors. The relative ease of measurement has ensured that much more self- and dopant diffusion data are available compared to the III–Vs. A further difference is that both cation and anion equilibrium vapor pressures are signif-

icant compared to the III–Vs, where the cation vapor pressures are negligible. Unless otherwise stated, the material in the following sections is drawn from the reviews by Shaw [6.6,76,77] and by Capper et al. [6.78].

6.9.1 Self-Diffusion

Where the anion self-diffusivity D_A has been measured as a function of the ambient anion or cation partial pressure in undoped material (ZnSe, CdS, CdSe, CdTe and $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$), a consistent pattern of behavior has emerged: in traversing the composition range from anion-rich to cation-rich, D_A is inversely proportional to the rising cation vapor pressure, P_C , until close to cation saturation, when D_A starts to increase with P_C . Strong donor doping in anion-rich CdS and CdSe had no effect on D_A . This evidence points to either a neutral anion interstitial or a neutral anion anti-site/anion vacancy complex as the diffusion mechanism over most of the composition range, changing to an anion vacancy mechanism as the cation-rich limit is approached.

The situation for cation self-diffusion proves to be more complicated due to the different variations of the cation self-diffusivity D_C with P_C across the compounds. In undoped ZnSe, ZnTe, CdTe and $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ (above $\approx 300^\circ\text{C}$), D_C is largely independent of P_C across the composition range. Such an independence excludes native point defect diffusion mechanisms and (excluding nondefect mechanisms) points to self-diffusion via neutral complexes such as a cation interstitial/cation vacancy or a cation vacancy/anion vacancy pair. Donor or acceptor doping increases D_C , indicating the involvement of ionized native defects or complexes. The Arrhenius parameters for Zn self-diffusion in undoped ZnSe above 760°C are $D_0 = 9.8\text{ cm}^2/\text{s}$ $Q = 3.0\text{ eV}$ and those for Hg in undoped $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ above 250°C are $D_0 = 3.8 \times 10^{-3}\text{ cm}^2/\text{s}$ and $Q = 1.22\text{ eV}$. In the case of undoped ZnS, CdS, CdSe and HgTe, D_C generally varies with P_C across the composition range. The simplest variations are found in CdSe and HgTe. In CdSe, D_C can be attributed to the parallel diffusion of singly (1+) and doubly (2+) ionized Cd self-interstitials. D_C in HgTe initially falls with P_C and then increases when crossing from anion-rich to cation-rich material, corresponding to diffusion by a singly ionized (1–) Hg vacancy and by a singly ionized (1+) Hg interstitial respectively. The behavior patterns in ZnS and CdS, however, present substantial problems in their interpretation: donor doping can also enhance D_C , point-

ing to the participation of an ionized native acceptor mechanism.

6.9.2 Chemical Self-Diffusion

Changes in the electrical conductivity or conductivity type caused by step changes to P_C in sample anneals have been used to characterize the change in level of nonstoichiometry through the chemical self-diffusivity, D_Δ , in CdS, CdTe and $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$. D_Δ obviously describes the diffusion of one or more ionized native defects, but in itself it does not identify the defect(s). In CdS and CdTe, D_Δ is attributed to the singly ionized (1+) and/or doubly ionized (2+) Cd interstitial; in CdTe, depending on the temperature, D_Δ exceeds D_C by a factor 10^5 to 10^6 . Modeling based on the simultaneous in-diffusion and out-diffusion of doubly ionized cation interstitials (2+) and vacancies (2–) gives a satisfactory quantitative account of type conversion ($p \rightarrow n$) in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ [6.7].

6.9.3 Dopant Diffusion

Although much information on dopant diffusion is available, it is mainly empirical and it is not uncommon for a dopant diffusivity to be independent of dopant concentration (as revealed by an *erfc* profile – a constant diffusivity for a given diffusion profile) under one set of conditions only to give profiles which cannot be characterized by single diffusivities when the conditions are changed. Equally, the variation of a dopant diffusivity with P_C may differ at different temperatures. A further difficulty when attempting to identify a diffusion mechanism is that the local electroneutrality condition is usually not known with any certainty due to significant concentrations of various ionized native defects. A good illustration of the problems encountered is provided by In diffusion in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$, where diffusion of the singly ionized (1–) pair $\text{In}_{\text{Hg}}\text{V}_{\text{Hg}}$ can account for some of the diffusion features. Some dopants, however, can present clear-cut diffusion properties which permit a well-defined interpretation. The diffusion of As in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ is one such case [6.79]. All of the observed features of $D(\text{As})$ are accounted for on the basis that: (i) As occupies both cation and anion lattice sites as singly ionized donors (1+) and acceptors (1–) respectively; (ii) only the ionized donor is mobile and diffuses by a vacancy mechanism on the cation sublattice; (iii) the diffusion sample is electrically intrinsic throughout, so the As concentration is always less than the intrinsic free carrier concentration.

6.9.4 Compositional Interdiffusion

Empirical information, based on bulk material, exists for CID in the following ternaries: (ZnCdHg)Te, (ZnCd)Se, (ZnCdHg)SeTe, (ZnCd)SSe, CdSeTe, ZnCdS, HgCdTe and CdMnTe. It might be expected that features evident in the binaries, such as donor doping enhancing the cation diffusivity but having no effect on that of the anion, and the anion diffusivity increasing (decreasing) with anion (cation) vapor pressure across most of the composition range, would continue to be seen. This means that in a ternary or quaternary system,

donor doping will enhance CID on the cation sublattice, but not on the anion sublattice, and annealing under a high (low) anion (cation) vapor pressure will enhance CID on the anion sublattice. This effect of the anion vapor pressure has been confirmed in CdSSe and CdSeTe and more recently in ZnSSe/ZnSe superlattices [6.80]. In (donor) doping has also been found to enhance the CID of the cations in CdMnTe [6.81], as has N (acceptor) doping in ZnMgSSe/ZnSSe superlattices [6.82]. The consequences of doping on CID in the II–VIs are obviously very similar to the III–V situation.

6.10 Conclusions

The first step in a diffusion investigation is to collect empirical data, which then leads to the second step where experiments can be designed to study the effects of the Fermi level (through the background doping level), of the ambient atmosphere (such as oxidizing, inert or vapor pressure of a system component) and of the sample structure (such as an MBE layer or a quantum well). The third step is to identify the diffusion mechanism and the associated defects using the experimental results in conjunction with the results from first-principles calculations of defect formation energies and their activation energies for diffusion. Clarification of the active processes involved can be gained

by numerical modeling (see Noda [6.83]). These data then provide the basis for the development of process simulators and defect engineering in which the concentrations and spatial distributions of host atoms, dopants and defects are organized according to requirement. Most progress towards achieving this ideal scenario has been made in Si and to a lesser extent in GaAs and $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$. The reality elsewhere is that the boundaries between the steps are blurred, with the third step often being undertaken with inadequate experimental information. Much work remains to be done in order to master our understanding of diffusion processes in semiconductors.

6.11 General Reading and References

General background material for diffusion in semiconductors can be found in Shaw [6.5], Tuck [6.1], Abdullaev and Dzhaferov [6.11] and Tan et al. [6.8]. More specific accounts are given by Fair [6.9] and Fahey et al. [6.84] for Si, by Frank et al. [6.3] for Si and Ge, by Tan and Gösele [6.43] for Si, Ge and GaAs, by Tuck [6.35] for the III–Vs and by Shaw [6.6, 77] for the II–VIs. *H in Semiconductors II* (1999) ed. by N. H. Nickel (*Semi-*

conductors and Semimetals, 61, Academic, San Diego) provides a recent account of H in semiconductors. The volumes in the EMIS Datareviews Series (IEE, Stevenage, UK) cover all of the important semiconductors. The series *Defects and Diffusion in Semiconductors* ed. by D. J. Fisher (Trans Tech., Brandrain 6, Switzerland) offers an annual and selective retrospective of recent literature.

References

- | | |
|--|---|
| <p>6.1 B. Tuck: <i>Introduction to Diffusion in Semiconductors</i> (Peregrinus, Stevenage 1974)</p> <p>6.2 A. Ural, P. B. Griffin, J. D. Plummer: <i>J. Appl. Phys.</i> 85, 6440 (1999)</p> | <p>6.3 W. Frank, U. Gösele, H. Mehrer, A. Seeger: In: <i>Diffusion in Crystalline Solids</i>, ed. by G. E. Murch, A. S. Nowick (Academic, Orlando 1984) Chapt.2</p> <p>6.4 D. Mathiot, J. C. Pfister: <i>J. Appl. Phys.</i> 66, 970 (1989)</p> |
|--|---|

- 6.5 D. Shaw: In: *Atomic Diffusion in Semiconductors*, ed. by D. Shaw (Plenum, London 1973) Chapt.1
- 6.6 D. Shaw: In: *Widegap II–VI Compounds for Optoelectronic Applications*, ed. by H. E. Ruda (Chapman and Hall, London 1992) Chapt.10
- 6.7 D. Shaw, P. Capper: *J. Mater. Sci. Mater. El.* **11**, 169 (2000)
- 6.8 T. Y. Tan, U. Gösele, S. Yu: *Crit. Rev. Sol. St. Mater. Sci.* **17**, 47 (1991)
- 6.9 R. B. Fair: In: *Impurity Doping Processes in Silicon*, ed. by F. F. Y. Wang (North-Holland, Amsterdam 1981) Chapt.7
- 6.10 S. M. Hu: *J. Appl. Phys.* **70**, R53 (1991)
- 6.11 G. B. Abdullaev, T. D. Dzhaferov: *Atomic Diffusion in Semiconductor Structures* (Harwood, Chur 1987)
- 6.12 M. Laudon, N. N. Carlson, M. P. Masquelier, M. S. Daw, W. Windl: *Appl. Phys. Lett.* **78**, 201 (2001)
- 6.13 K. Rajendran, W. Schoenmaker: *J. Appl. Phys.* **89**, 980 (2001)
- 6.14 H. Takeuchi, P. Ranada, V. Subramanian, T.-J. King: *Appl. Phys. Lett.* **80**, 3706 (2002)
- 6.15 S. C. Jain, W. Schoenmaker, R. Lindsay, P. A. Stolk, S. Decoutere, M. Willander, H. E. Maes: *J. Appl. Phys.* **91**, 8919 (2002)
- 6.16 L. Shao, J. Chen, J. Zhang, D. Tang, S. Patel, J. Liu, X. Wang, W.-K. Chu: *J. Appl. Phys.* **96**, 919 (2004)
- 6.17 Y. M. Haddara, J. C. Bravman: *Ann. Rev. Mater. Sci.* **28**, 185 (1998)
- 6.18 I. Lyubomirsky, V. Lyahovitskaya, D. Cahen: *Appl. Phys. Lett.* **70**, 613 (1997)
- 6.19 C. H. Chen, U. Gösele, T. Y. Tan: *Appl. Phys. A* **68**, 9, 19, 313 (1999)
- 6.20 P. N. Grillo, S. A. Stockman, J. W. Huang, H. Bracht, Y. L. Chang: *J. Appl. Phys.* **91**, 4891 (2002)
- 6.21 E. Chason, S. T. Picraux, J. M. Poate, J. O. Borland, M. I. Current, T. Diaz de la Rubia, D. J. Eaglesham, O. W. Holland, M. E. Law, C. W. Magee, J. W. Mayer, J. Melngailis, A. F. Tasch: *J. Appl. Phys.* **81**, 6513 (1997)
- 6.22 J. L. Melendez, J. Tregilgas, J. Dodge, C. R. Helms: *J. Electron. Mater.* **24**, 1219 (1995)
- 6.23 A. Borghesi, B. Pivac, A. Sassella, A. Stella: *J. Appl. Phys.* **77**, 4169 (1995)
- 6.24 K. F. Kelton, R. Falster, D. Gambaro, M. Olmo, M. Cornaro, P. F. Wei: *J. Appl. Phys.* **85**, 8097 (1999)
- 6.25 S. Solmi, E. Landi, F. Baruffaldi: *J. Appl. Phys.* **68**, 3250 (1990)
- 6.26 S. Solmi, D. Nobili: *J. Appl. Phys.* **83**, 2484 (1998)
- 6.27 B. Colombeau, N. E. B. Cowern: *Semicond. Sci. Technol.* **19**, 1339 (2004)
- 6.28 S. Mirabella, E. Bruno, F. Priolo, D. De Salvador, E. Napolitani, A. V. Drigo, A. Carnera: *Appl. Phys. Lett.* **83**, 680 (2003)
- 6.29 C. J. Ortiz, P. Pichler, T. Fühner, F. Cristiano, B. Colombeau, N. E. B. Cowern, A. Claverie: *J. Appl. Phys.* **96**, 4866 (2004)
- 6.30 D. Shaw: *Semicond. Sci. Technol.* **7**, 1230 (1992)
- 6.31 H. Puchner, S. Selberherr: *IEEE Trans. Electron. Dev.* **42**, 1750 (1995)
- 6.32 C. Poisson, A. Rolland, J. Bernardini, N. A. Stolwijk: *J. Appl. Phys.* **80**, 6179 (1996)
- 6.33 I. Kaur, Y. Mishin, W. Gust: *Fundamentals of Grain and Interphase Boundary Diffusion* (Wiley, Chichester 1995)
- 6.34 S. M. Hu: *J. Appl. Phys.* **43**, 2015 (1972)
- 6.35 B. Tuck: *Atomic Diffusion in III–V Semiconductors* (Adam Hilger, Bristol 1988)
- 6.36 L. S. Monastyrskii, B. S. Sokolovskii: *Sov. Phys. Semicond.* **16**, 1203 (1992)
- 6.37 E. A. Caridi, T. Y. Chang, K. W. Goossen, L. F. Eastman: *Appl. Phys. Lett.* **56**, 659 (1990)
- 6.38 A. Hangleiter, F. Hitzel, S. Lafmann, H. Rossow: *Appl. Phys. Lett.* **83**, 1169 (2003)
- 6.39 S. J. Rothman: In: *Diffusion in Crystalline Solids*, ed. by G. E. Murch, A. S. Nowick (Academic, Orlando 1984) Chapt.1
- 6.40 R. M. Fleming, D. B. McWhan, A. C. Gossard, W. Wiegmann, R. A. Logan: *J. Appl. Phys.* **51**, 357 (1980)
- 6.41 Y. L. Huang, Y. Ma, R. Job, W. R. Fahrner, E. Simeon, C. Claeys: *J. Appl. Phys.* **98**, 033511 (2005)
- 6.42 D. Mathiot: *Phys. Rev. B* **40**, 5867 (1989)
- 6.43 T. Y. Tan, U. Gösele: In: *Handbook of Semiconductor Technology*, Vol. 1, ed. by K. A. Jackson, W. Schröter (Wiley-VCH, Weinheim 2000) Chapt.5
- 6.44 M. Werner, H. Mehrer, H. D. Hochheimer: *Phys. Rev. B* **37**, 3930 (1985)
- 6.45 N. A. Stolwijk, W. Frank, J. Hölzl, S. J. Pearton, E. E. Haller: *J. Appl. Phys.* **57**, 5211 (1985)
- 6.46 A. Strohm, S. Matics, W. Frank: *Diffusion and Defect Forum* **194–199**, 629 (2001)
- 6.47 H. Bracht, E. E. Haller, R. Clark-Phelps: *Phys. Rev. Lett.* **81**, 393 (1998)
- 6.48 A. Ural, P. B. Griffin, J. D. Plummer: *Phys. Rev. Lett.* **83**, 3454 (1999)
- 6.49 A. Ural, P. B. Griffin, J. D. Plummer: *Appl. Phys. Lett.* **79**, 4328 (2001)
- 6.50 O. Krause, H. Ryssel, P. Pichler: *J. Appl. Phys.* **91**, 5645 (2002)
- 6.51 S. Solmi, A. Parisini, M. Bersani, D. Giubertoni, V. Soncini, G. Carnevale, A. Benvenuti, A. Marmiroli: *J. Appl. Phys.* **92**, 1361 (2002)
- 6.52 J. S. Christensen, H. H. Radamson, A. Yu. Kuznetsov, B. G. Svensson: *Appl. Phys. Lett.* **82**, 2254 (2003)
- 6.53 L. Lerner, N. A. Stolwijk: *Appl. Phys. Lett.* **86**, 011901 (2005)
- 6.54 N. R. Zangenberg, J. Fage-Pedersen, J. Lundsgaard Hansen, A. Nylandsted-Larsen: *Defect Diffus. Forum* **194–199**, 703 (2001)
- 6.55 N. R. Zangenberg, J. Fage-Pedersen, J. Lundsgaard Hansen, A. Nylandsted-Larsen: *J. Appl. Phys.* **94**, 3883 (2003)
- 6.56 A. D. N. Paine, A. F. W. Willoughby, M. Morooka, J. M. Bonar, P. Phillips, M. G. Dowsett, G. Cooke: *Defect Diffus. Forum* **143–147**, 1131 (1997)
- 6.57 J. S. Christensen, H. H. Radamson, A. Yu. Kuznetsov, B. G. Svensson: *J. Appl. Phys.* **94**, 6533 (2003)

- 6.58 D. B. Aubertine, P. C. McIntyre: J. Appl. Phys. **97**, 013531 (2005)
- 6.59 L. Torpo, M. Marlo, T. E. M. Staab, R. M. Nieminen: J. Phys. Condens. Matter **13**, 6203 (2001)
- 6.60 J. M. Lento, L. Torpo, T. E. M. Staab, R. M. Nieminen: J. Phys. Condens. Matter **16**, 1053 (2004)
- 6.61 M. Bockstedte, A. Mattausch, O. Pankratov: Phys. Rev. B **68**, 205201 (2003)
- 6.62 J. D. Hong, R. F. Davis, D. E. Newbury: J. Mater. Sci. **16**, 2485 (1981)
- 6.63 M. K. Linnarsson, M. S. Janson, J. Zhang, E. Janzen, B. G. Svensson: J. Appl. Phys. **95**, 8469 (2004)
- 6.64 Yu. A. Vodakov, G. A. Lomakina, E. N. Mokhov, V. G. Oding: Sov. Phys. Solid State **19**, 1647 (1977)
- 6.65 Yu. A. Vodakov, E. N. Mokhov: In: *Silicon Carbide – 1973*, ed. by R. C. Marshall, J. W. Faust Jr, C. E. Ryan (Univ. South Carolina Press, Columbia 1973) p. 508
- 6.66 I. O. Usov, A. A. Suvorova, Y. A. Kudriatsev, A. V. Suvorov: J. Appl. Phys. **96**, 4960 (2004)
- 6.67 N. Bagraev, A. Bouravleuv, A. Gippius, L. Klyachkin, A. Malyarenko: Defect Diffus. Forum **194–199**, 679 (2001)
- 6.68 S. Bai, R. P. Devaty, W. J. Choyke, U. Kaiser, G. Wagner, M. F. MacMillan: Appl. Phys. Lett. **83**, 3171 (2003)
- 6.69 N. A. Stolwijk, G. Bösker, J. Pöpping: Defect Diffus. Forum **194–199**, 687 (2001)
- 6.70 D. Shaw: Semicond. Sci. Technol. **18**, 627 (2003)
- 6.71 L. Wang, J. A. Wolk, L. Hsu, E. E. Haller, J. W. Erickson, M. Cardona, T. Ruf, J. P. Silveira, F. Brione: Appl. Phys. Lett. **70**, 1831 (1997)
- 6.72 J. C. Hu, M. D. Deal, J. D. Plummer: J. Appl. Phys. **78**, 1595 (1995)
- 6.73 J. Pöpping, N. A. Stolwijk, G. Bösker, C. Jäger, W. Jäger, U. Södervall: Defect Diffus. Forum **194–199**, 723 (2001)
- 6.74 I. Harrison: J. Mater. Sci. Mater. Electron. **4**, 1 (1993)
- 6.75 S. Govindaraju, J. M. Reifsnider, M. M. Oye, A. L. Holmes: J. Electron. Mater. **32**, 29 (2003)
- 6.76 D. Shaw: J. Cryst. Growth **86**, 778 (1988)
- 6.77 D. Shaw: J. Electron. Mater. **24**, 587 (1995)
- 6.78 P. Capper, C. D. Maxey, C. L. Jones, J. E. Gower, E. S. O’Keefe, D. Shaw: J. Electron. Mater. **28**, 637 (1999)
- 6.79 D. Shaw: Semicond. Sci. Technol. **15**, 911 (2000)
- 6.80 M. Kuttler, M. Grundmann, R. Heitz, U. W. Pohl, D. Bimberg, H. Stanzel, B. Hahn, W. Gebbhart: J. Cryst. Growth **159**, 514 (1994)
- 6.81 A. Barcz, G. Karczewski, T. Wojtowicz, J. Kossut: J. Cryst. Growth **159**, 980 (1996)
- 6.82 M. Strassburg, M. Kuttler, O. Stier, U. W. Pohl, D. Bimberg, M. Behringer, D. Hommel: J. Cryst. Growth **184–185**, 465 (1998)
- 6.83 T. Noda: J. Appl. Phys. **94**, 6396 (2003)
- 6.84 P. M. Fahey, P. B. Griffin, J. D. Plummer: J. Appl. Phys. **61**, 289 (1989)

7. Photoconductivity in Materials Research

Photoconductivity is the incremental change in the electrical conductivity of a substance upon illumination. Photoconductivity is especially apparent for semiconductors and insulators, which have low conductivity in the dark. Significant information can be derived on the distribution of electronic states in the material and on carrier generation and recombination processes from the dependence of the photoconductivity on factors such as the exciting photon energy, the intensity of the illumination or the ambient temperature. These results can in turn be used to investigate optical absorption coefficients or concentrations and distributions of defects in the material. Methods involving either steady state currents under constant illumination or transient methods involving pulsed excitation can be used to study the electronic density of states as well as the recombination. The transient time-of-flight technique also allows carrier drift mobilities to be determined.

7.1 Steady State Photoconductivity Methods	138
7.1.1 The Basic Single-Beam Experiment	138
7.1.2 The Constant Photocurrent Method (CPM)	141
7.1.3 Dual-Beam Photoconductivity (DBP)	141
7.1.4 Modulated Photoconductivity (MPC)	141
7.2 Transient Photoconductivity Experiments	142
7.2.1 Current Relaxation from the Steady State	143
7.2.2 Transient Photoconductivity (TPC)	143
7.2.3 Time-of-Flight Measurements (TOF)	144
7.2.4 Interrupted Field Time-of-Flight (IFTOF)	145
References	146

Photoconductivity has traditionally played a significant role in materials research, and most notably so in the study of covalently bonded semiconductors and insulators. Indeed, since it is the incremental conductivity generated by the absorption of (optical) photons, photoconductivity can be most clearly resolved in situations where the intrinsic dark conductivity of the material is low. This conductivity in the dark, leading to “dark current”, is due to the thermal equilibrium density of free carriers in the material and must be subtracted from any measured current in order to obtain the actual photocurrent. The basic processes that govern the magnitude of the photocurrent are the generation of free electrons and holes through the absorption of incident photons, the transport of those free carriers through the material under the influence of an electric field, and the recombination of the photoexcited electrons and holes. The study of any of those aspects as a function of the characteristics of the current-inducing illumination, as well as the study of their development upon changes in that illumi-

nation over time, will offer insights into the structure and electronic properties of the material under investigation. However, given the fact that three separate processes are involved in the production of a specific photocurrent, it follows that any analysis of experimental data in terms of system parameters will require a sufficiently comprehensive data set that will allow for differentiation between alternative interpretations. For instance, a low photocurrent may be the result of a low optical absorption coefficient at the given photon energy, but it may also be due to significant geminate recombination of the photogenerated electron-hole pairs, or it may reflect the formation of excitons. The combined use of different types of photoconductivity experiments is therefore often advisable, as is the combination of photoconductivity with related experiments such as photoluminescence or charge collection.

A wide variety of experimental techniques based on photoconductivity have come into general use over the years. They can be divided into two main groups, one

involving steady state photoconductivity (SSPC), where the focus is on stationary photocurrent levels, and a second one involving transient effects (TPC) where the time evolution of the photocurrent is studied. We will use this division in our survey of the various methods, but we should point out that SSPC can also be measured through ac excitation. The information that can be obtained about the material under investigation is generally not specific to either the SSPC or TPC method that is used, but will depend on the wider context of the

measurements. Recombination can be studied via TPC, but the temperature dependence of SSPC can also be used to identify different recombination mechanisms, while details of the electronic density of states (DOS) in the band gap of a semiconductor can be inferred either from the spectral response of the SSPC or from a proper analysis of TPC. Detailed discussions of the general principles of photoconductivity may be found in the standard monographs by Bube [7.1, 2], Ryvkin [7.3] and Rose [7.4].

7.1 Steady State Photoconductivity Methods

7.1.1 The Basic Single-Beam Experiment

The simplest photoconductivity experiment uses a constant monochromatic light source to generate equal excess densities of free electrons and holes, $\Delta n = \Delta p$, that lead to a change in the conductivity by

$$\Delta\sigma = \sigma_{ph} = e(\mu_n \Delta n + \mu_p \Delta p), \quad (7.1)$$

where e is the electronic charge and μ_n and μ_p are the electron and hole mobilities, respectively. The basic experimental arrangement is illustrated in Fig. 7.1a, where L and A are the length and the cross-sectional area of the sample and the photocurrent I_{ph} corresponds to $\sigma_{ph}AF$, where $F = V/L$ is the electric field applied. The end surfaces of the sample are covered by a metallic electrode. However, since materials of current interest are often used in thin film rather than bulk form, interdigitated electrodes of the type shown in Fig. 7.1b are frequently used in actual measurement geometries. In general, a fraction of the photogenerated carriers becomes immobilized by getting trapped at various defects such that not every part of Δn and Δp contributes equally to the photoconductivity in (7.1). The effect of such trapping on the photoconductivity is reflected in the use of values for the mobilities μ_n and μ_p that are lower – and not necessarily symmetrically lower – than the theoretical free-carrier mobility μ_0 . In fact, for

a significant number of materials with widespread practical applications, one of either the product $\mu_n \Delta n$ or the product $\mu_p \Delta p$ turns out to be much larger than the other because of strongly unequal carrier mobilities. For instance, the electron term dominates in intrinsic silicon,

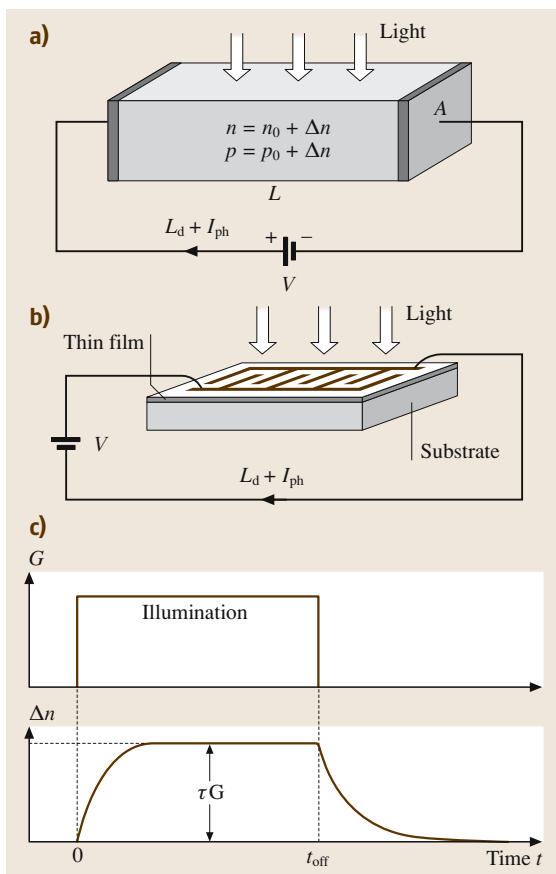


Fig. 7.1 (a) Basic arrangement for photoconductivity measurements, with V the applied voltage, L the sample length and A the cross-sectional area. I_d , n_0 and p_0 are the current and the carrier densities in the dark, and I_{ph} , Δn , Δp are the incremental values caused by the illumination. (b) Example of interdigitated electrode configuration for a thin film sample. (c) Schematic time development of the excess carrier concentration Δn in response to a period of illumination

while the photocurrent in chalcogenide glasses is carried by holes. In those instances, (7.1) effectively reduces to a one-carrier equation.

In the $\mu_n \Delta n$ or $\mu_p \Delta p$ products, the mobility μ_i is a material parameter that, in general, will depend on temperature and sample characteristics, while the excess carrier density $\Delta n = \Delta p$ is determined by a combination of material and external parameters. Phenomenologically, the excess density Δn can be written as the product $G\tau_i$, where G is the rate of generation of free electrons and holes per unit volume, and τ_i is the average lifetime of the excess carrier. Introducing these quantities into (7.1) leads to the form

$$\sigma_{ph} = eG(\mu_n \tau_n + \mu_p \tau_p), \quad (7.2)$$

which explicitly displays the mobility–lifetime products that are frequently used to characterize photoconductors. The relationship between the steady state values of Δn and G is illustrated in Fig. 7.1c, where the build-up and decay of Δn when the illumination is turned on and turned off are also shown. Those time-dependent aspects of photoconductivity will be addressed in a later section.

The generation rate G is defined by

$$G = \eta(I_0/h\nu)(1 - R)[1 - \exp(-\alpha d)]/d, \quad (7.3)$$

where η is the quantum efficiency of the generation process, I_0 is the incident illumination intensity (energy per unit time and unit area), $h\nu$ is the photon energy, R is the reflection coefficient of the sample, α is the optical absorption coefficient of the material, and d is the sample thickness. A quantum efficiency $\eta < 1$ signifies that, due to geminate recombination of the carriers or of exciton formation, not every absorbed photon generates a free electron and hole that will contribute to the photocurrent. The values of the parameters η , R and α depend, in general, on the wavelength of the illuminating light. Consequently, monochromatic illumination from a tunable light source can be used to obtain energy-resolved information about the sample, while illumination with white light will only offer a global average. Under many experimental circumstances, the condition $\alpha d \ll 1$ will hold over a significant energy range (when the sample thickness is small with respect to the optical absorption depth of the material). Equation (7.3) can then be simplified to

$$G \cong \eta(I_0/h\nu)(1 - R)\alpha. \quad (7.4)$$

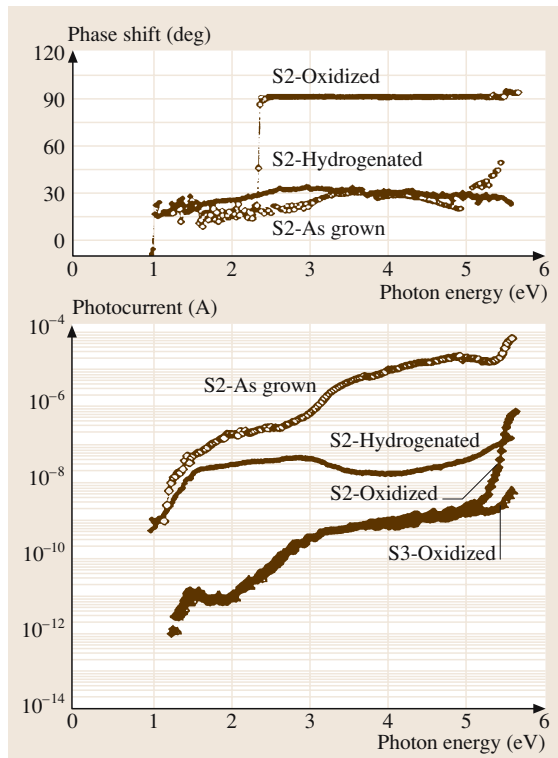
The free-carrier lifetimes of the excess electrons and holes, τ_n and τ_p , in (7.2) are governed by recombination with carriers of opposite sign. Assuming, for

simplicity, the frequently encountered case of photoconductivity dominated by one type of carrier (known as the majority carrier), and assuming electrons to be the majority carrier, the recombination rate can be written as $\tau_n^{-1} = b(p_0 + \Delta p)$, where b is a recombination constant, and p_0 and Δp are the equilibrium and excess minority carrier densities. It then follows that the photoconductivity

$$\begin{aligned} \sigma_{ph} \propto \Delta n &= G\tau_n = G/b(p_0 + \Delta p) \\ &= G/b(p_0 + \Delta n). \end{aligned} \quad (7.5)$$

Equation (7.5) indicates that a linear relationship $\sigma_{ph} \propto G$ holds for $\Delta n \ll p_0$ (a low excess carrier density), while high excitation levels with $\Delta n \gg p_0$ lead to $\sigma_{ph} \propto G^{1/2}$. These linear and quadratic recombination regimes are also referred to as mono- and bimolecular recombination. For a given light source and temperature, variations in G correspond to variations in the light intensity I_0 , and therefore $\sigma_{ph} \propto I_0^\gamma$ with $1/2 \leq \gamma \leq 1$. The value of γ itself will of course depend on the light intensity I_0 . However, I_0 is not the only factor that determines the value of γ : intermediate γ values may indicate a $\Delta n \approx p_0$ condition, but they may equally be caused by a distribution of recombination centers, as outlined below [7.4].

From a materials characterization point of view, SSPC offers the possibility of using the above equations to determine the absorption coefficient as a function of the energy of the incoming photons, and thus explore the electronic density of states around the band gap of a semiconductor. When single-crystalline samples of materials with sufficiently well-defined energy levels are studied, maxima corresponding to specific optical transitions may be seen in the photoconductivity spectra. A recent example, involving the split valence band of a p-CdIn₂Te₄ crystal, may be found in You et al. [7.5]. Another example is given in Fig. 7.2, where the spectral distribution of the photocurrent is shown for optical-quality diamond films prepared by chemical vapor deposition [7.6]. The rise in photocurrent around 5.5 eV corresponds to the optical gap of diamond, while the shoulders at ≈ 1.5 eV and ≈ 3.5 eV signal the presence of defect distributions in the gap. The data in Fig. 7.2 were obtained under ac conditions using chopped light and a lock-in amplifier. The changes in the observed phase shift can then also be used to locate the energies at which transitions to specific features of the density of states (DOS) become of importance. The use of ac excitation and lock-in detection has the added advantage of strongly reducing uncorrelated noise, but



care must be taken to ensure that the ac frequency remains lower than the response rate of the investigated system over the spectral range of interest.

The equilibrium free-carrier densities n_0 and p_0 , which play a role in the recombination pro-

Fig. 7.2 Room temperature ac photocurrent spectra, measured at 7 Hz, after various treatments of CVD diamond layers deposited at 920 °C (S2) and 820 °C (S3) (after [7.6])

cess, depend on the temperature through the approximate Fermi–Dirac occupation probability function $\exp[(E - E_F)/kT]$, thus making recombination a temperature-dependent process. In photoconductors, recombination is mediated by carrier traps in the bandgap. The presence of discrete trapping levels leads to thermally activated photocurrents, with the activation energy indicating the energetic positions of the traps. *Main and Owen* [7.8] and *Simmons and Taylor* [7.9] showed that the positive photocurrent activation energy in the monomolecular recombination regime corresponds to the distance above the Fermi level of a donor-like center, while a negative activation energy value in the bimolecular region refers to the energy position above the valence band edge of an acceptor-like center. Figure 7.3 illustrates this photocurrent behavior for amorphous As_2Se_3 [7.7]. The above pattern is characteristic of chalcogenide glasses, where the intrinsic charged defects with negative effective correlation energy act as recombination centers [7.10]. *SSPC* measurements can thus determine the recombination levels of those defects.

In highly photosensitive materials, such as selenium or hydrogenated amorphous silicon (a-Si:H), measurements in the monomolecular region are hindered by the problem of satisfying the $\Delta n \ll p_0$ condition. In addition, the *SSPC* temperature dependence in a-Si:H does not exhibit a definite activation energy due to the presence of a more distributed and complex set of traps that even induce regions of superlinear dependence on light intensity [7.11]. This illustrates that *SSPC* analysis is not necessarily straightforward.

Whenever the electronic density of states in the band gap of a photoconductor consists of a distribution of traps (as is the case in amorphous materials), a quasi-Fermi level $E_{qF} = E_F - kT \ln(1 + \Delta n/n)$, linked to the excess carrier density, can be defined. This quasi-Fermi level will – to a first approximation – correspond to the de-

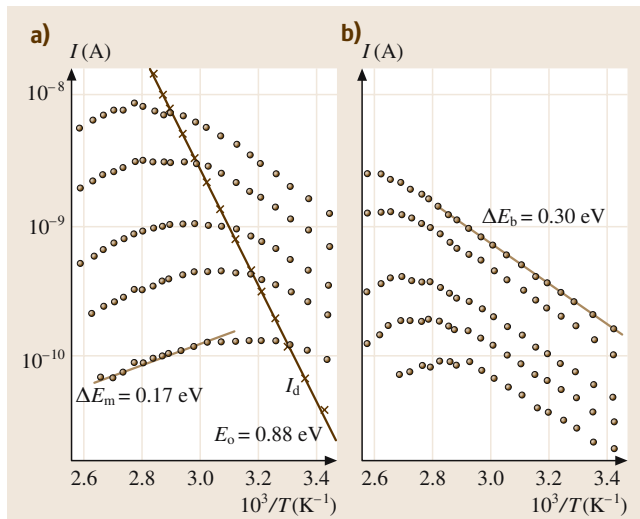


Fig. 7.3a,b Temperature dependence of the steady state dark and photocurrents in an a- As_2Se_3 bulk sample, illuminated at 1.55 eV with intensities of 0.84, 3.5, 9.8, 38 and 120×10^{12} photons/cm²s (a), and illuminated at 1.85 eV with intensities of 0.56, 1.7, 4.6, 27 and 77×10^{12} photons/cm²s (b). ΔE_m and ΔE_b represent the photocurrent activation energies in the monomolecular and bimolecular recombination regimes respectively, and E_σ is the activation energy of the dark current I_d (after [7.7])

marcation level that divides the DOS into a shallower part where carriers will be trapped and subsequently re-emitted and a deeper part where traps have become recombination centers. In other words, varying the light intensity influences both carrier generation and recombination rates. When several trapping centers with quite different characteristics are present in the photoconductor, shifts in the positions of the quasi-Fermi levels can then produce unexpected results. Instances of $\sigma_{ph} \propto I_0^\gamma$, with $\gamma > 1$ (as referred to above) will be observed for some materials, while combinations that actually produce *negative* photoconductivity, $\sigma_{ph} < 0$, have also been encountered [7.4].

7.1.2 The Constant Photocurrent Method (CPM)

The constant photocurrent method (CPM) has been used by Vaněček and coworkers [7.12, 13] to determine the optical absorption coefficient as a function of photon energy, $\alpha(E)$, via (7.2–7.4). In CPM, the photocurrent is kept constant by continually adjusting the light intensity I_0 while the photon energy is scanned across the spectrum. The constant photocurrent implies that the quasi-Fermi levels have immobile positions and thus that the free-carrier lifetime is a constant, τ . It then follows that in

$$\sigma_{ph} = e\mu\tau(I_0/h\nu)(1 - R)\eta\alpha \quad (7.6)$$

the product $(I_0/h\nu)\alpha$ will remain constant, and that α can be determined from it, provided that any energy dependencies for the parameters μ , R and η of (7.6) are negligible. The value at which the photocurrent is fixed can be chosen freely, but will in practice be dictated by the low-absorption region of the sample. However, since even low-level photocurrents can still be measured with high precision, the method is especially useful at low values of optical absorption where standard transmission measurements lose their accuracy.

In ‘absolute’ CPM, the optical transmission through the film is measured at the same time as the photocurrent, and the data from the two measurements are combined in order to remove optical interference fringes from the data and to fix the value of the proportionality constant [7.12]. The experimental arrangement used in such absolute CPM measurements is shown schematically in Fig. 7.4. The CPM experiment can be operated with either dc or ac illumination, but the absorption spectra retrieved will not be identical. AC illumination can be obtained using a mechanical chopper (as suggested in Fig. 7.4), but also, for instance, from an ac-driven light-emitting

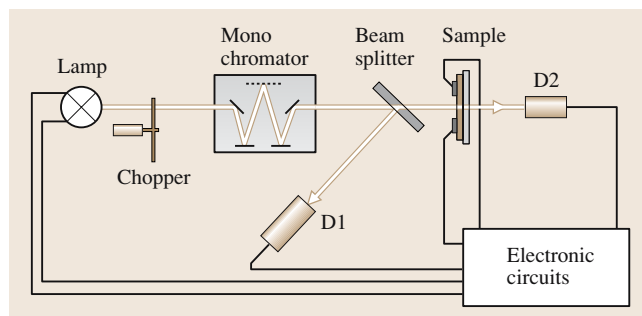


Fig. 7.4 Schematic diagram of an ‘absolute’ CPM set-up. Photodetector D1 is used to regulate the intensity of the lamp, while detector D2 measures the transmitted light (after [7.12])

diode. Main et al. [7.14] showed that, in the dc mode, transitions involving initially unoccupied DOS levels raise the absorption above the value that is seen with the ac technique. Systematic comparison of *dc* and *ac* results allows us, therefore, to distinguish between occupied states below the operative Fermi level and unoccupied ones above it. In cases where the quantum efficiency of carrier generation η can be taken as unity, CPM gives $\alpha(E)$ directly as $1/I_0$, and this method is widely used, for example for hydrogenated amorphous silicon. However, for materials such as chalcogenide glasses or organic semiconductors where η itself is energy-dependent, it is only the product $\eta\alpha$ that is readily obtained.

7.1.3 Dual-Beam Photoconductivity (DBP)

Like the CPM discussed above, the dual-beam photoconductivity (DBP) technique is used to determine the sub-bandgap optical absorption in a photoconductor. A constant, uniformly absorbed illumination I_0 is used to establish a constant excess carrier density in the material, and hence a constant free-carrier lifetime τ . The chopped signal $I'(E)$ of a low-intensity, tunable light source is added to this background to generate variations in photoconductivity $\delta\sigma_{ph}(E)$. Synchronous lock-in detection of the small ac signal then provides the information needed to deduce $\alpha(E)$. By carrying out measurements at different values of the background illumination intensity, DBP allows the photoconductor absorption to be tested for changing quasi-Fermi level positions. Changes in the resolved $\alpha(E)$ curves can then be used to obtain information on the DOS distribution in the sample. An example of this use of DBP can be found in Günes et al. [7.15], where differences in absorption between annealed and light-soaked hydrogenated amorphous silicon samples are studied.

7.1.4 Modulated Photoconductivity (MPC)

The experimental technique that has become known as modulated photoconductivity (MPC) is used to determine the energetic distribution of states in the bandgap of a photoconductor by analyzing the phase shift between ac photoexcitation and the ensuing ac photocurrent as a function of the modulation frequency of the light [7.17, 18]. Figure 7.5 shows the essential parts of an MPC set-up, and illustrates the phase difference between the illumination and the photocurrent. Two modulation frequency ranges with distinct characteristics are identified. In the high-frequency region, from a few Hz up to the kHz range, the signal is dominated by carrier release from traps, with a release rate that matches the modulation frequency. The usual assumption, that the release probability decreases exponentially with the trap depth according to $r \propto \exp(-E/kT)$, gives the link between the measured phase shift and the DOS of the material. The relationship between the two is expressed by

$$g(E) \propto \sin(\Phi)/I_{ac}, E = k_B T \ln(v_0/\omega), \quad (7.7)$$

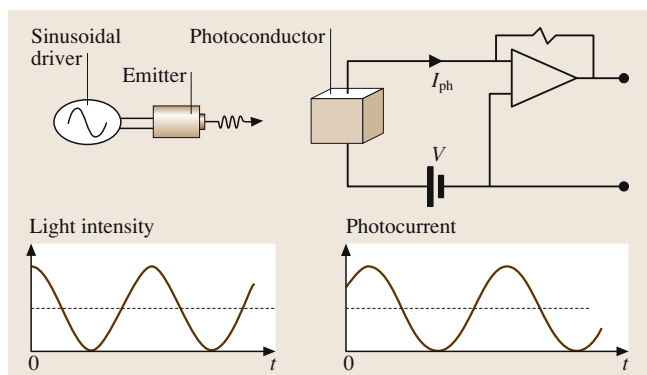


Fig. 7.5 Schematic diagram of an MPC set-up (upper frames), and of the phase relationship between the exciting light intensity and the resulting photocurrent (lower frames)

where Φ and I_{ac} are the phase and intensity of the ac photocurrent, k_B is the Boltzmann constant, T the temperature, v_0 the attempt-to-escape frequency and ω the modulation frequency. At the low-frequency end, recombination and trapping in deep states determine the phase shifts and the DOS varies according to $\tan(\Phi)/\omega$. The transition between the two regions is tied to the position of the quasi-Fermi levels and can, therefore, be shifted by changing the illumination intensity. MPC works best with photoconductors where one carrier type dominates the current, and therefore only one side of the bandgap needs be taken into account in the analysis. Examples of MPC-determined DOS profiles are given in Fig. 7.6 [7.16]. The figure shows a profile for the conduction band side of the bandgap of an as-deposited polymorphous silicon sample, as well as those for the sample following light soaking and after subsequent annealing.

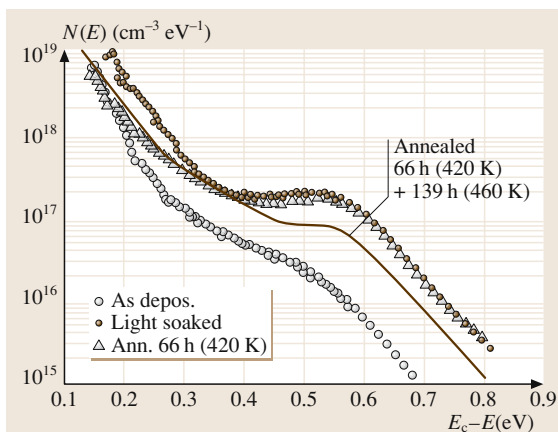


Fig. 7.6 DOS below the CB edge of a polymorphous silicon sample deposited at 423 K and measured (by MPC) as-deposited, after light soaking, and after two stages of subsequent annealing (after [7.16])

7.2 Transient Photoconductivity Experiments

The study of transient aspects of photoconductivity can relate to either the build-up or the relaxation of steady state photocurrents, or to a material's response to pulsed excitation. While the SSPC turn-on transient reflects the interplay between generation and recombination of carriers (an interplay that often leads to a current overshoot at high excitation levels), the SSPC relaxation upon turn-off only involves recombination and is therefore

easier to analyze. Nevertheless, a simple exponential decay of the photocurrent, as sketched in Fig. 7.1c, will only be observed when a unique recombination path is followed, a situation that is the exception rather than the rule. Transient photoconductivity (TPC) caused by pulsed excitation is generally simpler to analyze. Indeed, whereas a quasi-equilibrium distribution of trapped photogenerated carriers will build up or be present in the

photoconductor's bandgap under **SSPC**, the **TPC** experiments can be analyzed against the background of the thermal equilibrium distribution of carriers in the material.

7.2.1 Current Relaxation from the Steady State

Upon termination of steady state illumination, the generation term drops out of the rate equation that describes the nonequilibrium carrier distribution, but the carrier density itself and the operative recombination process are not altered. Consequently, the initial photocurrent decay will be governed by whatever recombination mode existed under **SSPC** conditions. Spectroscopic analysis of the relaxation current in terms of the distribution of states in the bandgap can be readily achieved in the case of monomolecular recombination [7.19], with the product of photocurrent and time being proportional to the **DOS**:

$$I_{\text{ph}}(t)t \propto g(E), \quad E = k_B T \ln(v_0 t). \quad (7.8)$$

In (7.8), $k_B T$ is the Boltzmann energy and v_0 is the attempt-to-escape frequency. When, on the other hand, bimolecular recombination dominates, the link between the current and the distribution of recombination centers is much less direct and spectroscopic analysis is difficult. Unfortunately, bimolecular recombination is dominant in good photoconductors.

In spite of the above, relaxation of the steady state current has often been used to obtain a first-order estimate for free-carrier lifetimes, even when this had to be done on a purely phenomenological basis due to a lack of sufficient information on the recombination mechanisms involved. An exponential fit to the initial part of the decay is then often used to make the estimate. In cases where more than one – sometimes vastly different – recombination mechanisms are operative, this initial decay does not necessarily represent the most significant proportion of carriers. This is certainly the case whenever so-called persistent photoconductivity is observed; one of the relaxation times involved is then longer than the observation time.

7.2.2 Transient Photoconductivity (TPC)

In the standard transient photoconductivity (**TPC**) experiment, free carriers are excited into the transport band at time $t = 0$ by a short light pulse. They are then moved along by the electric field until their eventual disappearance through recombination, but before this happens

they will have been immobilized a number of times by various traps that are present in the material. Since the carrier distributions are in thermal equilibrium at the start of the experiment, both the trapping sites for electrons above the Fermi level and the hole trapping sites below E_F are empty, such that the newly created carriers are not excluded from any of those trapping sites. Given that carrier release from a trap is a thermally activated process with the trap depth being the activation energy, deeper traps immobilize carriers for longer times and lead to lower values for the transient current. As shallower states release trapped carriers sooner, retrapping of those carriers will lead to increased occupation of the deeper states and further reduction of the current level. To allow this thermalization of the excited carriers to run its full course until recombination sets in, the experiments are traditionally carried out in the so-called secondary photocurrent mode, whereby the sample is supplied with ohmic electrical contacts and carrier loss is by recombination only. Coplanar electrode geometries (gap cells) are mostly used. Expressions that link the transient current to the distribution of localized states can be derived [7.20], but they are difficult to invert in the general case. Nevertheless, as long as recombination can be neglected, the relationship $g(E) \propto [I(t)t]^{-1}$ can be used as a first-order estimate.

For the special case of an exponential **DOS**, the solution is straightforward: a $g(E) \propto \exp(-E/E_0)$ distribution of trapping levels leads to a power law for the transient current $I(t) \propto t^{-(1-\alpha)}$ with $\alpha = k_B T/E_0$. In other words, the width of the exponential distribution E_0 can be deduced from the slope of the power law decay of the current. Essentially exponential distributions were found to dominate the valence band tail of equilibrated amorphous As_2Se_3 samples over a wide energy range [7.7], but no other examples have emerged.

An elegant way to circumvent the difficulties posed by a time domain analysis of the transient current is to transpose the current decay into the frequency domain by a Fourier transform [7.21]. Since the **TPC** current decay is the photoconductor's response to an impulse excitation, its Fourier transform gives the frequency response $I(\omega)$ of that photoconductor. In fact, this $I(\omega)$ corresponds to the photocurrent intensity I_{ac} as used in the **MPC** method, and the same procedures can thus be used to extract the information on density and energy distribution of localized states in the band gap. Not just Fourier transform but also Laplace transform techniques have been applied to the conversion of **TPC** signals into **DOS** information. A comparison and discussion of the results may be found in [7.22]. Examples of Fourier

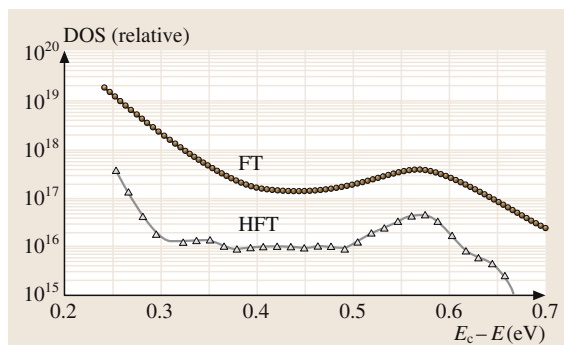


Fig. 7.7 DOS below the conduction band edge in a-Si:H, obtained through Fourier transforms of the transient photocurrent; HFT: the high-resolution analysis of [7.22], FT: the earlier analysis according to [7.21]

transform TPC analysis, as originally proposed and as developed since, are shown in Fig. 7.7 for an a-Si:H sample. Whereas the energy range that can be probed is limited in MPC by the frequency range of the lock-in amplifier, it is the smallest resolution time of the detection system that limits the range in the case of TPC, the latter one being generally more advantageous.

7.2.3 Time-of-Flight Measurements (TOF)

The time-of-flight (TOF) experiment, originally designed to determine the drift mobility of free carriers in high-mobility materials, has been highly successfully adapted to low-mobility materials such as organic or amorphous semiconductors [7.24], where it has been used for drift mobility measurements but also as an alternative TPC technique to study the energy distribution of localized states. While majority carriers will dominate photocurrents in traditional TPC, TOF allows independent measurements with majority and minority carriers, and thus independent examinations of the valence band side and conduction band side of the band gap.

For TOF measurements, the sample consists of a layer of the photoconducting material sandwiched between two electrodes that are blocking carrier injection into the sample. At least one of the electrodes must be semitransparent to permit the photoexcitation of free carriers in the material just beyond the illuminated contact by a strongly absorbed light flash. Depending on the polarity of the electric field applied across the sample, either electrons or holes will then be drifted through the sample. At their arrival at the back contact, the current will drop since the blocking contact ensures that only the primary photocurrent is measured. From the transit time

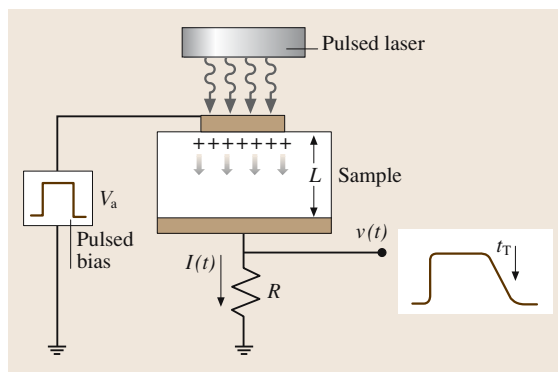


Fig. 7.8 TOF measurement set-up shown for the case of holes being drifted through the length L of the sample by a positive applied voltage. Choosing the resistance R that generates an output voltage to be low minimizes RC distortion of the signal at short times; choosing it to be high enhances the detectability of weak signals at the expense of time resolution

t_T (the time needed for the charge sheet to cross the sample), the drift mobility μ_d can be calculated according to $\mu_d = L/t_T F$, where L is the sample length and F the applied electric field. The essential elements of a TOF

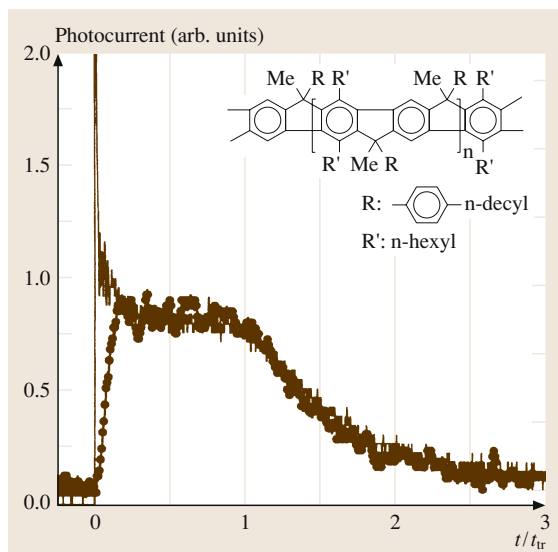


Fig. 7.9 Time-of-flight transients measured at 243 K in methyl-substituted ladder-type poly(para)phenylene (MeLPPP) with 60 kV/cm (line) and 300 kV/cm (●) applied, and normalized to a transit time set to 90% of the pre-transit current. The inset shows the chemical structure of MeLPPP (after [7.23])

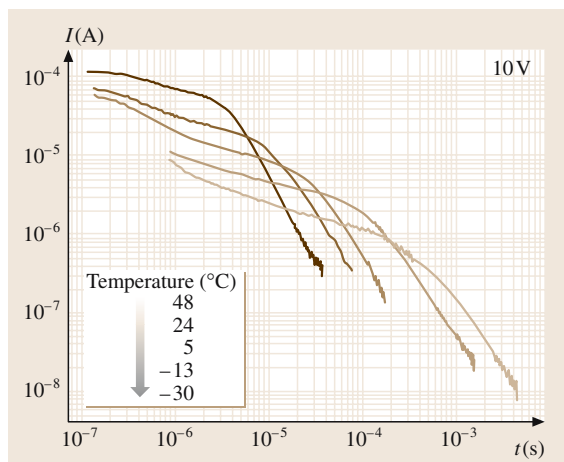


Fig. 7.10 Example of TOF hole transients measured at several temperatures, as indicated, with 10 V applied across a 5.6 μm -thick a-Si:H sample grown in an expanding thermal plasma at 0.85 nm/s and 250 $^{\circ}\text{C}$ substrate temperature, and sandwiched between Mo contacts (after [7.25])

measuring circuit are displayed in Fig. 7.8. The transit time can be measured directly on the current trace, in which case it is variously defined as the time at which the current has dropped by values ranging from 10% to 50% (the latter one being most commonly used), or it can be obtained by integrating the current and using the time at which the collected charge saturates. Obtaining a true value of μ_d requires that the field F be uniform and constant during the carrier transit, which means that F should only be applied a short time before the optical excitation and that the transit time should be short with respect to the dielectric relaxation time in the material. Figure 7.9 shows TOF transients in a conjugated polymer whereby a 10% drop is used to define the transit time.

In materials with a wide distribution of localized gap states, as is generally the case in disordered photoconductors, the drifting charge package spreads out along the length of the sample, and a representative transit time can only be discerned as a change of slope in a double-logarithmic plot of current versus time. The curves in Fig. 7.10 (from [7.25]) illustrate such behavior. Measurements at different temperatures and applied fields are then needed to ascertain that the observed feature marks an actual carrier transit rather than deep trapping of the photogenerated charge. In the materials that exhibit this anomalously dispersive transport, the pre-transit current will have the characteristics of the TPC described in the previous section,

and the information about the distribution of gap states $g(E)$ that is contained in the current transient can be extracted in the same ways. Both pre-transit current transients and measured drift mobility values have been employed in the past to estimate the DOS in the band tails of disordered semiconductors. In the latter case, specific $g(E)$ functions are explored through trap-controlled transport modeling to reproduce the experimental dependence of μ_d on the temperature and the electric field. This technique has since been replaced by the more direct procedures described in preceding sections.

At times longer than the TOF transit time, a steeper current decay testifies to the fact that carriers are leaving the sample. The post-transit current that is then observed is increasingly due to the emission of carriers that were trapped in states deep in the bandgap. Provided that the conditions are such that the probability of subsequent deep retrapping of the same carriers is negligible, a proper analysis of these post-transit TOF current transients permits the elucidation of the distribution of localized states deeper in the gap [7.27] with, as in (7.8), $g(E) \propto I(t)t$ expressing the correspondence.

7.2.4 Interrupted Field Time-of-Flight (IFTOF)

The interrupted field time-of-flight (IFTOF) experiment differs from the time-of-flight experiment described

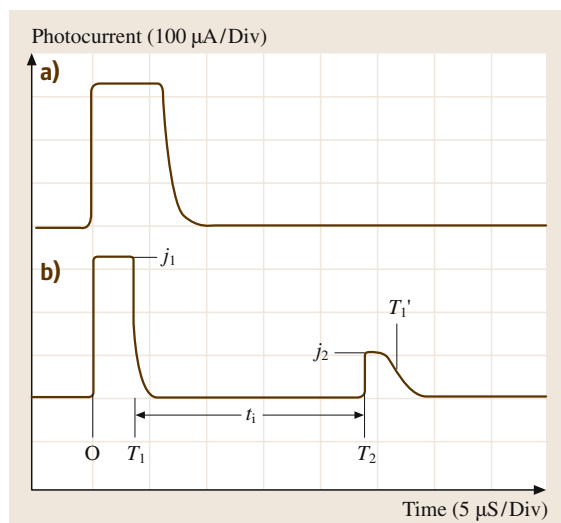


Fig. 7.11a,b Comparison of current traces in TOF (a) and IFTOF (b) experiments. The applied electric field is turned off in case (b) for a length of time t_1 (after [7.26])

in the previous section in that the applied field that drives the photogenerated carrier packet through the sample is turned off for some period of time before the carriers have completed their transit. As illustrated in Fig. 7.11, a lower current intensity is measured when the field is turned on again, signalling that some of the drifting carriers have become immobilized in deep traps [7.28]. By studying the drop in current as a function of the interruption time t_i , the deep-trapping lifetime of the carriers can be evaluated. Recombination can be routinely neglected in TOF experiments since only one type of carrier drifts through the sample, but by charging a sample with carriers of one polarity before performing an IFTOF experiment that drifts carriers of the opposite polarity through

the sample, recombination parameters can be studied too [7.29].

Another interesting method for studying the recombination process is – just like IFTOF – based on a simple modification of the TOF experiment: after generating free carriers through one contact and drifting the slower type of carrier into the sample, a second light pulse through the other contact sends a sheet of oppositely charged carriers towards the first one. The two carrier packages will cross and some electrons and holes will recombine during that crossing, thereby affecting the observed current levels and providing a way to study the recombination process. An elegant example of the application of this technique to amorphous selenium can be found in Haugen and Kasap [7.30].

References

- 7.1 R. H. Bube: *Photoconductivity of Solids* (Wiley, New York 1960)
- 7.2 R. H. Bube: *Photoelectronic Properties of Semiconductors* (Cambridge Univ. Press, Cambridge 1992)
- 7.3 S. M. Ryvkin: *Photoelectric Effects in Semiconductors* (Consultants Bureau, New York 1964)
- 7.4 A. Rose: *Concepts in Photoconductivity and Allied Problems* (Krieger, Huntington 1978)
- 7.5 S. H. You, K. J. Hong, T. S. Jeong, C. J. Youn, J. S. Park, D. C. Shin, J. D. Moon: *J. Appl. Phys.* **95**, 4042 (2004)
- 7.6 M. Nesládek, L. M. Stals, A. Stesmans, K. Iakoubovskii, G. J. Adriaenssens, J. Rosa, M. Vaněček: *Appl. Phys. Lett.* **72**, 3306 (1998)
- 7.7 G. J. Adriaenssens: *Philos. Mag. B* **62**, 79 (1990) and references therein
- 7.8 C. Main, A. E. Owen: In: *Electronic and Structural Properties of Amorphous Semiconductors*, ed. by P. G. Le Comber, J. Mort (Academic, London 1973) p. 527
- 7.9 J. G. Simmons, G. W. Taylor: *J. Phys. C* **7**, 3051 (1974)
- 7.10 G. J. Adriaenssens, N. Qamhieh: *J. Mater. Sci. Mater. El.* **14**, 605 (2003)
- 7.11 H. Fritzsche, B.-G. Yoon, D.-Z. Chi, M. Q. Tran: *J. Non-Cryst. Solids* **141**, 123 (1992)
- 7.12 M. Vaněček, J. Kočka, A. Poruba, A. Fejfar: *J. Appl. Phys.* **78**, 6203 (1995)
- 7.13 M. Vaněček, J. Kočka, J. Stuchlík, A. Tríska: *Solid State Commun.* **39**, 1199 (1981)
- 7.14 C. Main, S. Reynolds, I. Zrinščak, A. Merazga: *Mater. Res. Soc. Symp. Proc.* **808**, 103 (2004)
- 7.15 M. Günes, C. Wronski, T. J. McMahon: *J. Appl. Phys.* **76**, 2260 (1994)
- 7.16 C. Longeaud, D. Roy, O. Saadane: *Phys. Rev. B* **65**, 85206 (2002)
- 7.17 H. Oheda: *J. Appl. Phys.* **52**, 6693 (1981)
- 7.18 R. Brüggemann, C. Main, J. Berkin, S. Reynolds: *Philos. Mag. B* **62**, 29 (1990)
- 7.19 M. S. Iovu, I. A. Vasiliev, E. P. Colomeico, E. V. Emelianova, V. I. Arkhipov, G. J. Adriaenssens: *J. Phys. Condens. Mat.* **16**, 2949 (2004)
- 7.20 A. I. Rudenko, V. I. Arkhipov: *Philos. Mag. B* **45**, 209 (1982)
- 7.21 C. Main, R. Brüggemann, D. P. Webb, S. Reynolds: *Solid State Commun.* **83**, 401 (1992)
- 7.22 C. Main: *J. Non-Cryst. Solids* **299**, 525 (2002)
- 7.23 D. Hertel, A. Ochse, V. I. Arkhipov, H. Bässler: *J. Imag. Sci. Technol.* **43**, 220 (1999)
- 7.24 W. E. Spear: *J. Non-Cryst. Solids* **1**, 197 (1969)
- 7.25 M. Brinza, E. V. Emelianova, G. J. Adriaenssens: *Phys. Rev. B* **71**, 115209 (2005)
- 7.26 S. Kasap, B. Polishuk, D. Dodds, S. Yannacopoulos: *J. Non-Cryst. Solids* **114**, 106 (1989)
- 7.27 G. F. Seynhaeve, R. P. Barclay, G. J. Adriaenssens, J. M. Marshall: *Phys. Rev. B* **39**, 10196 (1989)
- 7.28 S. Kasap, B. Polishuk, D. Dodds: *Rev. Sci. Instrum.* **61**, 2080 (1990)
- 7.29 S. Kasap, B. Fogal, M. Z. Kabir, R. E. Johanson, S. K. O'Leary: *Appl. Phys. Lett.* **84**, 1991 (2004)
- 7.30 C. Haugen, S. O. Kasap: *Philos. Mag. B* **71**, 91 (1995)

8. Electronic Properties of Semiconductor Interfaces

In this chapter we investigate the electronic properties of semiconductor interfaces. Semiconductor devices contain metal–semiconductor, insulator–semiconductor, insulator–metal and/or semiconductor–semiconductor interfaces. The electronic properties of these interfaces determine the characteristics of the device. The band structure lineup at all these interfaces is determined by one unifying concept, the continuum of interface-induced gap states (IFIGS). These intrinsic interface states are the wavefunction tails of electron states that overlap the fundamental band gap of a semiconductor at the interface; in other words they are caused by the quantum-mechanical tunneling effect. IFIGS theory quantitatively explains the experimental barrier heights of well-characterized metal–semiconductor or Schottky contacts as well as the valence-band offsets of semiconductor–semiconductor interfaces or

8.1	Experimental Database	149
8.1.1	Barrier Heights of Laterally Homogeneous Schottky Contacts .	149
8.1.2	Band Offsets of Semiconductor Heterostructures	152
8.2	IFIGS-and-Electronegativity Theory	153
8.3	Comparison of Experiment and Theory ..	155
8.3.1	Barrier Heights of Schottky Contacts	155
8.3.2	Band Offsets of Semiconductor Heterostructures	156
8.3.3	Band-Structure Lineup at Insulator Interfaces	158
8.4	Final Remarks	159
	References	159

semiconductor heterostructures. Insulators are viewed as semiconductors with wide band gaps.

In his pioneering article entitled *Semiconductor Theory of the Blocking Layer*, Schottky [8.1] finally explained the rectifying properties of metal–semiconductor contacts, which had first been described by Braun [8.2], as being due to a depletion of the majority carriers on the semiconductor side of the interface. This new depletion-layer concept immediately triggered a search for a physical explanation of the barrier heights observed in metal–semiconductor interfaces, or Schottky contacts as they are also called in order to honor Schottky’s many basic contributions to this field.

The early Schottky–Mott rule [8.3,4] proposed that n-type (p-type) barrier heights were equal to the difference between the work function of the metal and the electron affinity (ionization energy) of the semiconductor. A plot of the experimental barrier heights of various metal–selenium rectifiers versus the work functions of the corresponding metals did indeed reveal a linear correlation, but the slope parameter was much smaller than unity [8.4]. To resolve the failure of

the very simple and therefore attractive Schottky–Mott rule, Bardeen [8.5] proposed that electronic interface states in the semiconductor band gap play an essential role in the charge balance at metal–semiconductor interfaces.

Heine [8.6] considered the quantum-mechanical tunneling effect at metal–semiconductor interfaces and noted that *for energies in the semiconductor band gap, the volume states of the metal have tails in the semiconductor*. Tejedor and Flores [8.7] applied this same idea to semiconductor heterostructures where, for energies in the band-edge discontinuities, the volume states of one semiconductor tunnel into the other. The continua of interface-induced gap states (IFIGS), as these evanescent states were later called, are an intrinsic property of semiconductors and they are the *fundamental* physical mechanism that determines the band-structure lineup at both metal–semiconductor contacts and semiconductor heterostructures: in other words, at all semiconductor interfaces. Insulator interfaces are also included in this, since insulators may be described as wide-gap semi-

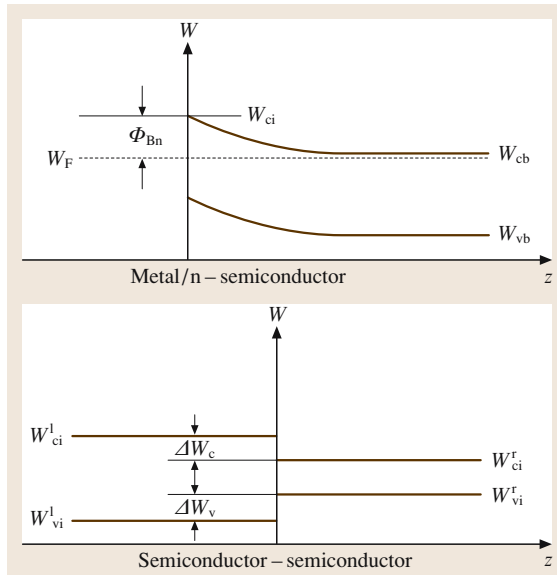


Fig. 8.1 Schematic energy-band diagrams of metal–semiconductor contacts and semiconductor heterostructures. W_F : Fermi level; Φ_{Bn} : barrier height; W_v and W_c : valence-band maximum and conduction-band minimum, respectively; ΔW_v and ΔW_c : valence- and conduction-band offset, respectively; i and b: values at the interface and in the bulk, respectively; r and l: right and left side, respectively

conductors. Figure 8.1 shows schematic band diagrams of an n-type Schottky contact and a semiconductor heterostructure.

The **IFIGS** continua derive from both the valence- and the conduction-band states of the semiconductor. The energy at which their predominant character changes from valence-band-like to conduction-band-like is called their branch point. The position of the Fermi level relative to this branch point then determines the sign and the amount of the net charge in the **IFIGS**. Hence, the **IFIGS** give rise to intrinsic interface dipoles. Both the barrier heights of Schottky contacts and the band offsets of heterostructures thus divide up into a zero-charge-transfer term and an electric-dipole contribution.

From a more chemical point of view, these interface dipoles may be attributed to the partial ionic character of the covalent bonds between atoms right at the interface. Generalizing *Pauling's* [8.8] electronegativity concept,

the difference in the electronegativities of the atoms involved in the interfacial bonds also describes the charge transfer at semiconductor interfaces. Combining the physical **IFIGS** and the chemical electronegativity concept, the electric-dipole contributions of Schottky barrier heights as well as those of heterostructure band offsets vary proportional to the difference in the electronegativities of the metal and the semiconductor and of the two semiconductors, respectively. The electronegativities of the Group IV elemental and the IV–IV, III–V, and II–VI compound semiconductors are almost equal, since the elements that constitute these semiconductors are all placed in the middle of the Periodic Table. Hence, the **IFIGS** dipole terms of the respective semiconductor heterostructures will be small and may be neglected [8.9]. The valence-band offsets of nonpolar, of lattice-matched and of metamorphic heterostructures should thus equal the difference between the branch-point energies of the semiconductors in contact.

The theoreticians appreciated *Heine's* **IFIGS** concept at once. The initial reluctance of most experimentalists was motivated by the observation that the predictions of the **IFIGS** theory only marked upper limits for the barrier heights observed with *real* Schottky contacts [8.10]. *Schmitsdorf* et al. [8.11] finally resolved this dilemma. They found a linear decrease in the effective barrier height with increasing ideality factors for their Ag/n-Si(111) diodes. Such behavior has been observed for all of the Schottky contacts investigated so far. *Schmitsdorf* et al. attributed this correlation to patches of decreased barrier heights and lateral dimensions smaller than the depletion layer width [8.12]. Consequently, they extrapolated their plots of effective barrier height versus ideality factor to the ideality factor determined by the image-force or Schottky effect [8.13] alone; in this way, they obtained the barrier heights of the laterally homogeneous contacts. The barrier heights of laterally uniform contacts can also be determined from capacitance–voltage measurements (C/V) and by applying ballistic-electron-emission microscopy (**BEEM**) and internal photoemission yield spectroscopy (**IPEYS**). The I/V , C/V , **BEEM**, and **IPEYS** data agree within the margins of experimental error.

Mönch [8.14] found that the barrier heights of laterally homogeneous Schottky contacts as well as the experimentally observed valence band offsets of semiconductor heterostructures agree excellently with the predictions of the **IFIGS**-and-electronegativity theory.

8.1 Experimental Database

8.1.1 Barrier Heights of Laterally Homogeneous Schottky Contacts

I/V Characteristics

The current transport in real Schottky contacts occurs via thermionic emission over the barrier provided the doping level of the semiconductor is not too high [8.15]. For doping levels larger than approximately 10^{18} per cm^3 , the depletion layer becomes so narrow that tunnel or field emission through the depletion layer prevails. The current–voltage characteristics then become ohmic rather than rectifying.

For thermionic emission over the barrier, the current–voltage characteristics may be written as (see, for example, [8.14])

$$I_{\text{te}} = AA_{\text{R}}^* T^2 \exp\left(-\Phi_{\text{Bn}}^{\text{eff}}/k_{\text{B}}T\right) \exp(e_0 V_{\text{c}}/nk_{\text{B}}T) \times [1 - \exp(-e_0 V_{\text{c}}/k_{\text{B}}T)], \quad (8.1)$$

where A is the diode area, A_{R}^* is the effective Richardson constant of the semiconductor, and k_{B} , T , and e_0 are Boltzmann's constant, the temperature, and the electronic charge, respectively. The effective Richardson constant is defined as

$$A_{\text{R}}^* = \frac{4\pi e_0 k_{\text{B}} m_n^*}{h^3} = A_{\text{R}} \frac{m_n^*}{m_0}, \quad (8.2)$$

where $A_{\text{R}} = 120 \text{ A cm}^{-2} \text{ K}^{-2}$ is the Richardson constant for thermionic emission of nearly free electrons into vacuum, h is Planck's constant, and m_0 and m_n^* are the vacuum and the effective conduction-band mass of electrons, respectively. The externally applied bias V_{a} divides up into a voltage drop V_{c} across the depletion layer of the Schottky contact and an IR drop at the series resistance R_{s} of the diode, so that $V_{\text{c}} = V_{\text{a}} - IR_{\text{s}}$. For *ideal* (intimate, abrupt, defect-free, and, above all, laterally homogeneous) Schottky contacts, the effective zero-bias barrier height $\Phi_{\text{Bn}}^{\text{eff}}$ equals the difference $\Phi_{\text{Bn}}^{\text{hom}} - \delta\Phi_{\text{if}}^0$ between the homogeneous barrier height and the zero-bias image-force lowering (see [8.14])

$$\delta\Phi_{\text{if}}^0 = e_0 \left[\frac{2e_0^2 N_{\text{d}}}{(4\pi)^2 \varepsilon_{\infty}^2 \varepsilon_{\text{b}} \varepsilon_0^3} \left(e_0 |V_{\text{i}}^0| - k_{\text{B}}T \right) \right]^{1/4}, \quad (8.3)$$

where N_{d} is the donor density, $e_0 |V_{\text{i}}^0|$ is the zero-bias band bending, ε_{∞} and ε_{b} are the optical and the bulk dielectric constant, respectively, and ε_0 is the permittivity

of vacuum. The ideality factor n describes the voltage dependence of the barrier height and is defined by

$$1 - 1/n = \partial\Phi_{\text{Bn}}^{\text{eff}}/\partial e_0 V_{\text{c}}. \quad (8.4)$$

For *real* diodes, the ideality factors n are generally found to be larger than the ideality factor

$$n_{\text{if}} = \left(1 - \frac{\delta\Phi_{\text{if}}^0}{4e_0 |V_{\text{i}}^0|} \right)^{-1}, \quad (8.5)$$

which is determined by the image-force effect only.

The effective barrier heights and the ideality factors of real Schottky diodes fabricated under experimentally identical conditions vary from one specimen to the next. However, the variations of both quantities are correlated, and the $\Phi_{\text{Bn}}^{\text{eff}}$ values become smaller as the ideality factors increase. As an example, Fig. 8.2 displays $\Phi_{\text{Bn}}^{\text{eff}}$ versus n data for Ag/n-Si(111) contacts with $(1 \times 1)^{\text{i}}$ -unreconstructed and $(7 \times 7)^{\text{i}}$ -reconstructed interfaces [8.11]. The dashed and dash-dotted lines are the linear least-squares fits to the data points. The linear dependence of the effective barrier height on the ideality factor may be written as

$$\Phi_{\text{Bn}}^{\text{eff}} = \Phi_{\text{Bn}}^{\text{nif}} - \varphi_{\text{p}}(n - n_{\text{if}}), \quad (8.6)$$

where $\Phi_{\text{Bn}}^{\text{nif}}$ is the barrier height at the ideality factor n_{if} . Several conclusions may be drawn from this relation. First, the $\Phi_{\text{Bn}}^{\text{eff}} - n$ correlation shows that more than

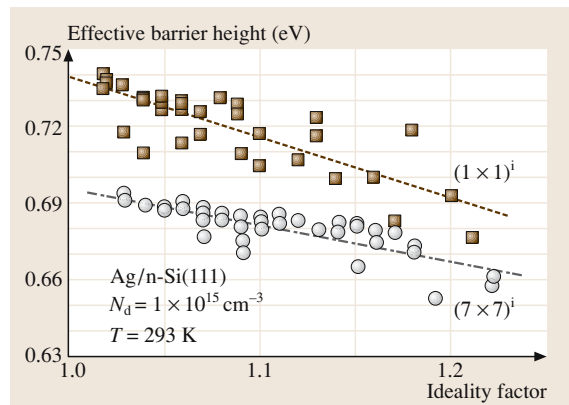


Fig. 8.2 Effective barrier heights versus ideality factors determined from I/V characteristics of Ag/n-Si(111)- $(7 \times 7)^{\text{i}}$ and $-(1 \times 1)^{\text{i}}$ contacts at room temperature. The *dashed* and *dash-dotted* lines are the linear least-squares fits to the data. After [8.11]

one physical mechanism determines the barrier heights of *real* Schottky contacts. Second, the extrapolation of $\Phi_{\text{Bn}}^{\text{eff}}$ versus n curves to n_{if} removes all mechanisms that cause a larger bias dependence of the barrier height than the image-force effect itself from consideration. Third, the extrapolated barrier heights $\Phi_{\text{Bn}}^{\text{nif}}$ are equal to the zero-bias barrier height $\Phi_{\text{Bn}}^{\text{hom}} - \delta\Phi_{\text{if}}^0$ of the laterally homogeneous contact.

The laterally homogeneous barrier heights obtained from $\Phi_{\text{Bn}}^{\text{eff}}$ versus n curves to n_{if} are not necessarily characteristic of the corresponding *ideal* contacts. This is illustrated by the two data sets displayed in Fig. 8.2, which differ in the interface structures of the respective diodes. Quite generally, structural rearrangements such as the $(7 \times 7)^i$ reconstruction are connected with a redistribution of the valence charge. The bonds in perfectly ordered bulk silicon, the example considered here, are purely covalent, and so reconstructions are accompanied by electric $\text{Si}^{+\Delta q} - \text{Si}^{-\Delta q}$ dipoles. The $\text{Si}(111)-(7 \times 7)$ reconstruction is characterized by a stacking fault in one half of its unit mesh [8.16]. *Schmitsdorf* et al. [8.11] quantitatively explained the experimentally observed reduction in the laterally homogeneous barrier height of the $(7 \times 7)^i$ with regard to the $(1 \times 1)^i$ diodes by the electric dipole associated with the stacking fault of the $\text{Si}(111)-7 \times 7$ reconstruction.

Patches of reduced barrier height with lateral dimensions smaller than the depletion layer width that are embedded in large areas of laterally homogeneous barrier height is the only known model that explains a lowering of effective barrier heights with increasing ideality factors. In their phenomenological studies of such patchy Schottky contacts, *Freeouf* et al. [8.12] found that the potential distribution exhibits a saddle point in front of such nanometer-size patches of reduced barrier height. Figure 8.4 explains this behavior. The saddle-point barrier height strongly depends on the voltage drop V_c across the depletion layer. *Freeouf* et al. simulated the current transport in such patchy Schottky contacts and found a reduction in the effective barrier height and a correlated increase in the ideality factor as they reduced the lateral dimensions of the patches. However, they overlooked the fact that the barrier heights of the laterally homogeneous contacts may be obtained from $\Phi_{\text{Bn}}^{\text{eff}}$ versus n plots, by extrapolating to n_{if} .

C/V Characteristics

Both the space charge and the width of the depletion layers at metal–semiconductor contacts vary as a function of the externally applied voltage. The space-charge the-

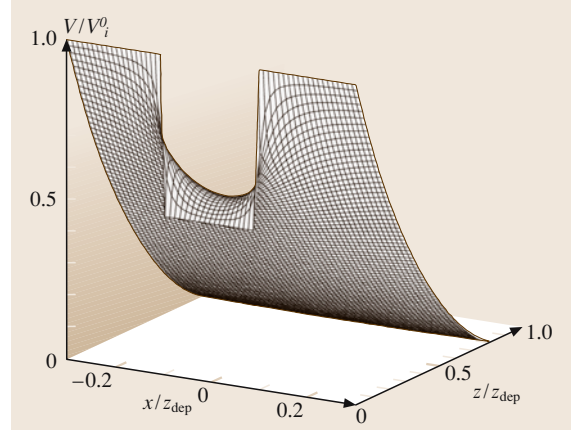


Fig. 8.3 Calculated potential distribution underneath and around a patch of reduced interface potential embedded in a region of larger interface band-bending. The lateral dimension and the interface potential reduction of the patch are set to two tenths of the depletion layer width z_{dep} and one half of the interface potential of the surrounding region

ory gives the variation in the depletion layer capacitance per unit area as (see [8.14])

$$C_{\text{dep}} = \{e_0^2 \epsilon_b \epsilon_0 N_d / 2 [e_0 (|V_i^0| - V_c) - k_B T]\}^{1/2}. \quad (8.7)$$

The current through a Schottky diode biased in the reverse direction is small, so the *IR* drop due to the series resistance of the diode may be neglected. Consequently, the extrapolated intercepts on the abscissa of $1/C_{\text{dep}}^2$ versus V_a plots give the band bending $e_0 |V_i^0|$ at the interface, and together with the energy distance $W_n = W_F - W_{\text{cb}}$ from the Fermi level to the conduction band minimum in the bulk, one obtains the flat-band barrier height $\Phi_{\text{Bn}}^{\text{fb}} \equiv \Phi_{\text{Bn}}^{\text{hom}} = e_0 |V_i^0| + W_n$ which equals the laterally homogeneous barrier height of the contact.

As an example, Fig. 8.4 displays the flat-band barrier heights of the same $\text{Ag}/n\text{-Si}(111)$ diodes that are discussed in Fig. 8.2. The dashed and dash-dotted lines are the Gaussian least-squares fits to the data from the diodes with $(1 \times 1)^i$ and $(7 \times 7)^i$ interface structures, respectively. Within the margins of experimental error the peak C/V values agree with the laterally homogeneous barrier heights obtained from the extrapolations of the I/V data shown in Fig. 8.2. These data clearly demonstrate that barrier heights characteristic of laterally homogeneous Schottky contacts can be only obtained from I/V or C/V data from many diodes fabricated under identical conditions rather than from a single diode. However, the

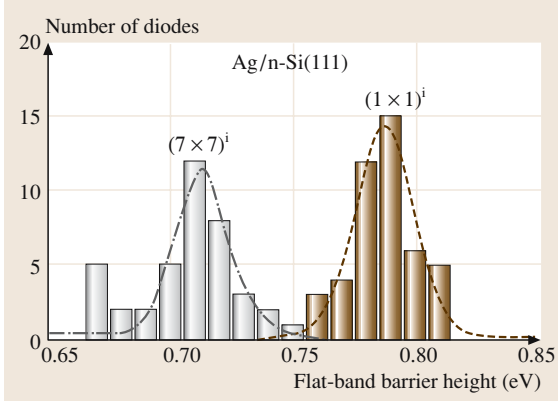


Fig. 8.4 Histograms of flat-band barrier heights determined from C/V characteristics of Ag/n-Si(111)-(7×7)ⁱ and -(1×1)ⁱ contacts at room temperature. The data were obtained with the same diodes discussed in Fig. 8.2. The *dashed* and *dash-dotted lines* are the Gaussian least-squares fits to the data. After [8.11]

effective barrier heights and the ideality factors vary as a function of the diode temperature. Hence, effective barrier heights and ideality factors evaluated from the I/V characteristics for one and the same diode recorded at different temperatures are also suitable for determining the corresponding laterally homogeneous barrier height (see [8.14]).

Ballistic-Electron-Emission Microscopy

In ballistic-electron-emission microscopy (BEEM) [8.18], a tip injects almost monoenergetic electrons into the metal film of a Schottky diode. These tunnel-injected electrons reach the semiconductor as ballistic electrons provided that they lose no energy on their way through the metal. Hence, the collector current I_{coll} is expected to set in when the ballistic electrons surpass the metal-semiconductor barrier; in other words, if the voltage V_{tip} applied between tip and metal film exceeds the *local* potential barrier $\Phi_{\text{Bn}}^{\text{loc}}(z)/e_0$. Bell and Kaiser [8.19] derived the square law

$$I_{\text{coll}}(z) = R^* I_{\text{tip}} \left[e_0 V_{\text{tip}} - \Phi_{\text{Bn}}^{\text{loc}}(z) \right]^2 \quad (8.8)$$

for the BEEM $I_{\text{coll}}/V_{\text{tip}}$ characteristics, where I_{tip} is the injected tunnel current. BEEM measures *local* barrier heights; most specifically, the saddle-point barrier heights in front of nanometer-sized patches rather than their lower barrier heights right at the interface.

BEEM is the experimental tool for measuring spatial variations in the barrier height on the nanometer-scale.

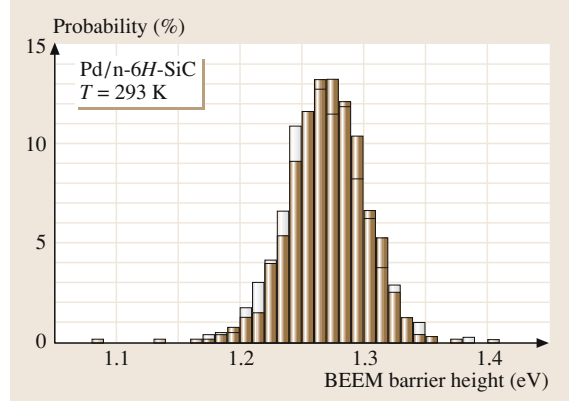


Fig. 8.5 Histograms of local BEEM barrier heights of two Pd/n-6H-SiC(0001) diodes with ideality factors of 1.06 (gray solid bars) and 1.49 (empty bars). The data were obtained by fitting the square law (8.8) to 800 BEEM $I_{\text{coll}}/V_{\text{tip}}$ spectra each. Data from Im et al. [8.17]

The local barrier heights are determined by fitting relation (8.8) to measured $I_{\text{coll}}/V_{\text{tip}}$ characteristics recorded at successive tip positions along lateral line scans. Figure 8.5 displays histograms of the local BEEM barrier heights of two Pd/n-6H-SiC(0001) diodes [8.17]. The diodes differ in their ideality factors, 1.06 and 1.49, which are close to and much larger, respectively, than the value $n_{\text{if}} = 1.01$ determined solely by the image-force effect. Obviously, the *nanometer-scale* BEEM histograms of the two diodes are identical although their *macroscopic* ideality factors and therefore their patchiness differ. Two important conclusions were drawn from these findings. First, these data suggest the existence of two different types of patches, intrinsic and extrinsic ones. The intrinsic patches might be correlated with the random distributions of the ionized donors and acceptors which cause nanometer-scale lateral fluctuations in the interface potential. A few gross interface defects of extrinsic origin, which escape BEEM observations, are then responsible for the variations in the ideality factors. Second, Gaussian least-squares fits to the histograms of the local BEEM barrier heights yield peak barrier heights of 1.27 ± 0.03 eV. Within the margins of experimental error, this value agrees with the laterally homogeneous value of 1.24 ± 0.09 eV which was obtained by extrapolation of the linear least-squares fit to a $\Phi_{\text{Bn}}^{\text{eff}}$ versus n plot to n_{if} . The nanometer-scale BEEM histograms and the macroscopic I/V characteristics thus provide identical barrier heights of laterally homogeneous Schottky contacts.

Internal Photoemission Yield Spectroscopy

Metal-semiconductor contacts show a photoelectric response to optical radiation with photon energies smaller than the width of the bulk band gap. This effect is caused by photoexcitation of electrons from the metal over the interfacial barrier into the conduction band of the semiconductor. Experimentally, the internal photoemission yield, which is defined as the ratio of the photoinjected electron flux across the barrier into the semiconductor to the flux of the electrons excited in the metal, is measured as a function of the energy of the incident photons. Consequently, this technique is called internal photoemission yield spectroscopy (IPEYS). *Cohen et al. [8.21]* derived that the internal photoemission yield varies as a function of the photon energy $\hbar\omega$ as

$$Y(\hbar\omega) \propto \left(\hbar\omega - \Phi_{\text{Bn}}^{\text{IPEYS}} \right)^2 / \hbar\omega. \quad (8.9)$$

Patches only cover a small portion of the metal-semiconductor interface, so the threshold energy $\Phi_{\text{Bn}}^{\text{IPEYS}}$ will equal the barrier height $\Phi_{\text{Bn}}^{\text{hom}}$ of the laterally homogeneous part of the contact minus the zero-bias image-force lowering $\delta\Phi_{\text{if}}^0$.

In Fig. 8.6, experimental $[Y(\hbar\omega) \cdot \hbar\omega]^{1/2}$ data for a Pt/p-Si(001) diode [8.20] are plotted versus the energy of the exciting photons. The dashed line is the linear least-squares fit to the data. The deviation of the experimental $[Y(\hbar\omega) \cdot \hbar\omega]^{1/2}$ data towards larger values slightly below and above the threshold is caused by the shape of the Fermi-Dirac distribution function at finite temperatures and by the existence of patches with barrier heights smaller and larger than $\Phi_{\text{Bn}}^{\text{hom}}$.

8.1.2 Band Offsets of Semiconductor Heterostructures

Semiconductors generally grow layer-by-layer, at least initially. Hence, core-level photoemission spectroscopy (PES) is a very reliable tool and the one most widely used to determine the band-structure lineup at semiconductor heterostructures. The valence-band offset may be obtained from the energy positions of core-level lines in X-ray photoelectron spectra recorded with bulk samples of the semiconductors in contact and with the interface itself [8.22]. Since the escape depths of the photoelectrons are on the order of just 2 nm, one of the two semiconductors must be sufficiently thin. This condition is easily met when heterostructures are grown by molecular beam epitaxy (MBE) and PE spectra are recorded during growth interrupts.

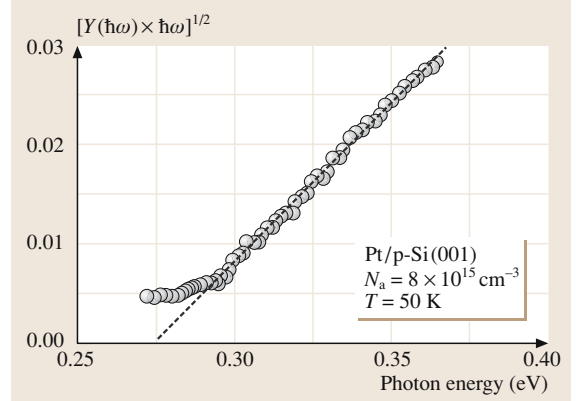


Fig. 8.6 Spectral dependence of the internal photoemission yield $\sqrt{Y(\hbar\omega) \cdot \hbar\omega}$ of a Pt/p-Si(001) diode versus the photon energy of the exciting light. The dashed line is the linear least-squares fit to the data for photon energies larger than 0.3 eV. Data from *Turan et al. [8.20]*

The valence-band discontinuity is then given by (see Fig. 8.7)

$$\begin{aligned} \Delta W_v = W_{\text{vir}} - W_{\text{vil}} = & W_i(n_r l_r) - W_i(n_l l_l) \\ & + [W_{\text{vbr}} - W_b(n_r l_r)] - [W_{\text{vbl}} - W_b(n_l l_l)], \end{aligned} \quad (8.10)$$

where $n_r l_r$ and $n_l l_l$ denote the core levels of the semiconductors on the right (r) and the left (l) side of the interface, respectively. The subscripts i and b characterize interface and bulk properties, respectively. The

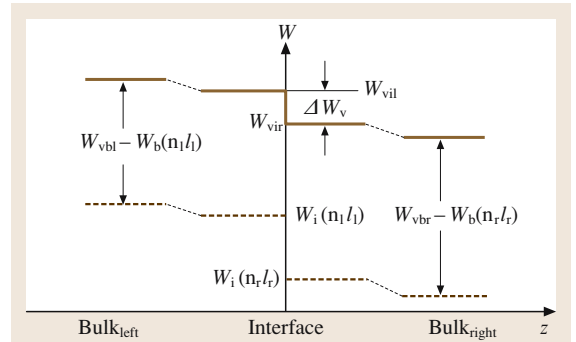


Fig. 8.7 Schematic energy band diagram at semiconductor heterostructures. W_{vb} and W_{vi} are the valence-band maxima and $W_b(nl)$ and $W_i(nl)$ are the core levels in the bulk and at the interface, respectively. The subscripts l and r denote the semiconductors on the right and the on the left side of the interface. ΔW_v is the valence-band offset. The thin dashed lines account for possible band-bending from space-charge layers

energy difference $W_i(n_r l_r) - W_i(n_l l_l)$ between the core levels of the two semiconductors at the interface is determined from energy distribution curves of photoelectrons recorded during MBE growth of the heterostructure. The energy positions $W_{vbr} - W_b(n_r l_r)$ and $W_{vbl} - W_b(n_l l_l)$ of the core levels relative to the valence-band maxima

in the bulk of the two semiconductors are evaluated separately.

Another widely used technique for determining band offsets in heterostructures is internal photoemission yield spectroscopy. The procedure for evaluating the IPEYS signals is the same as described in Sect. 8.1.1.

8.2 IFIGS- and Electronegativity Theory

Because of the quantum-mechanical tunneling effect, the wavefunctions of bulk electrons decay exponentially into vacuum at surfaces or, more generally speaking, at solid–vacuum interfaces. A similar behavior occurs at interfaces between two solids [8.6, 7]. In energy regions of Schottky contacts and semiconductor heterostructures where occupied band states overlap a band gap, the wavefunctions of these electrons will tail across the interface. The only difference to solid–vacuum interfaces is that the wavefunction tails oscillate at solid–solid interfaces. Figure 8.8 schematically explains the tailing effects at surfaces and semiconductor interfaces. For the band-structure lineup at semiconductor interfaces, only the tailing states within the gap between the top valence and the lowest conduction band are of any real importance since the energy position of the Fermi level determines their charging state. These wavefunction tails or interface-induced gap states (IFIGS) derive from the continuum of the virtual gap states (ViGS) of the complex semiconductor band structure. Hence, the IFIGS are an intrinsic property of the semiconductor.

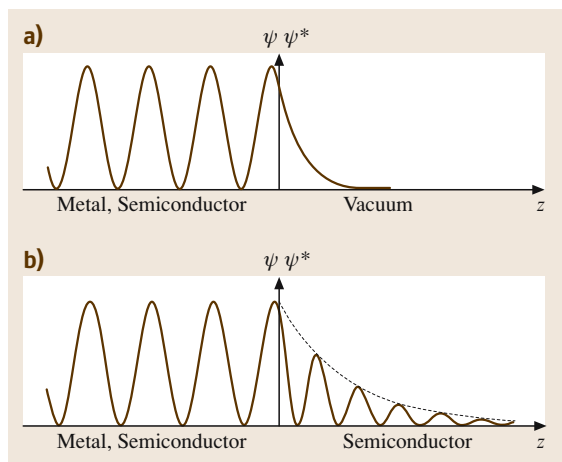


Fig. 8.8a,b Wavefunctions at clean surfaces (a) and at metal–semiconductor and semiconductor–semiconductor interfaces (b) (schematically)

The IFIGS are made up of valence-band and conduction-band states of the semiconductor. Their net charge depends on the energy position of the Fermi level relative to their branch point, where their character changes from predominantly donor- or valence band-like to mostly acceptor- or conduction band-like. The band-structure lineup at semiconductor interfaces is thus described by a zero-charge-transfer term and an electric dipole contribution.

In a more chemical approach, the charge transfer at semiconductor interfaces may be related to the partly ionic character of the covalent bonds at interfaces. *Pauling* [8.8] described the ionicity of single bonds in diatomic molecules by the difference between the electronegativities of the atoms involved. The binding energies of core-level electrons are known to depend on the chemical environment of the atoms or, in other words, on the ionicity of their chemical bonds. Figure 8.9 displays experimentally observed chemical shifts for Si(2p) and Ge(3d) core levels induced by metal adatoms on silicon and germanium surfaces.

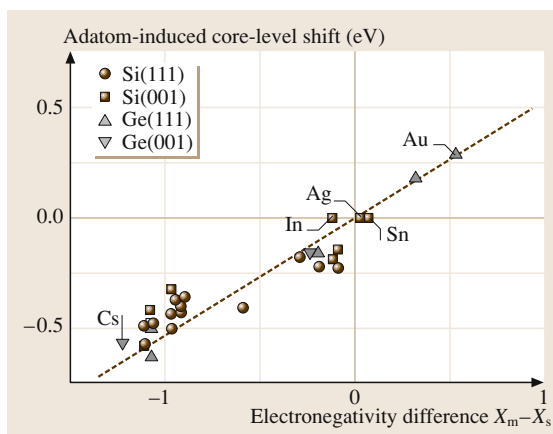


Fig. 8.9 Chemical shifts of Si(2p) and Ge(3d) core levels induced by metal adatoms on silicon and germanium surfaces, respectively, as a function of the difference $X_m - X_s$ in the metal and the semiconductor electronegativities in Pauling units. After [8.14]

on silicon and germanium surfaces as a function of the difference $X_m - X_s$ between the Pauling atomic electronegativity of the metal and that of the semiconductor atoms. The covalent bonds between metal and substrate atoms still persist at metal–semiconductor interfaces, as ab-initio calculations [8.23] have demonstrated for the example of Al/GaAs(110) contacts. The pronounced linear correlation of the data displayed in Fig. 8.9 thus justifies the application of *Pauling's* electronegativity concept to semiconductor interfaces.

The combination of the physical **IFIGS** and the chemical electronegativity concept yields the barrier heights of ideal p-type Schottky contacts and the valence-band offsets of ideal semiconductor heterostructures as

$$\Phi_{\text{Bp}} = \Phi_{\text{bp}}^{\text{p}} - S_X(X_m - X_s) \quad (8.11)$$

and

$$\Delta W_v = \Phi_{\text{bpr}}^{\text{p}} - \Phi_{\text{bpl}}^{\text{p}} + D_X(X_{\text{sr}} - X_{\text{sl}}), \quad (8.12)$$

respectively, where $\Phi_{\text{bp}}^{\text{p}} = W_{\text{bp}} - W_v(\Gamma)$ is the energy distance from the valence-band maximum to the branch point of the **IFIGS** or the p-type branch-point energy. It has the physical meaning of a zero-charge-transfer barrier height. The slope parameters S_X and D_X are explained at the end of this section.

The **IFIGS** derive from the virtual gap states of the complex band structure of the semiconductor. Their branch point is an average property of the semiconductor. *Tersoff* [8.24, 27] calculated the branch-point energies $\Phi_{\text{bp}}^{\text{p}}$ of Si, Ge, and 13 of the III–V and II–VI compound semiconductors. He used a linearized augmented plane-wave method and the local density approximation. Such extensive computations may be avoided. *Mönch* [8.28] applied *Baldereschi's* concept [8.29] of mean-value k -points to calculate the branch-point energies of zincblende-structure compound semiconductors. He first demonstrated that the quasi-particle band gaps of diamond, silicon, germanium, 3C–SiC, GaAs and CdS at the mean-value k -point equal their average or dielectric band gaps [8.30]

$$W_{\text{dg}} = \hbar\omega_p / \sqrt{\epsilon_{\infty} - 1}, \quad (8.13)$$

where $\hbar\omega_p$ is the plasmon energy of the bulk valence electrons. *Mönch* then used *Tersoff's* $\Phi_{\text{bp}}^{\text{p}}$ values, calculated the energy dispersion $W_v(\Gamma) - W_v(k_{\text{mv}})$ of the top-most valence band in the empirical tight-binding approximation (ETB), and plotted the resulting branch-point energies $W_{\text{bp}} - W_v(k_{\text{mv}}) = \Phi_{\text{bp}}^{\text{p}} + [W_v(\Gamma) - W_v(k_{\text{mv}})]_{\text{ETB}}$ at the mean-value k -point k_{mv} versus the widths of the dielectric band gaps W_{dg} . The linear least-squares

Table 8.1 Optical dielectric constants, widths of the dielectric band gap, and branch-point energies of diamond-, zincblende- and chalcopyrite-structure semiconductors and of some insulators

Semiconductor	ϵ_{∞}	$W_{\text{dg}}(\text{eV})$	$\Phi_{\text{bp}}^{\text{p}}(\text{eV})$
C	5.70	14.40	1.77
Si	11.90	5.04	0.36 ^a
Ge	16.20	4.02	0.18 ^a
3C–SiC	6.38	9.84	1.44
3C–AlN	4.84	11.92	2.97
AlP	7.54	6.45	1.13
AlAs	8.16	5.81	0.92
AlSb	10.24	4.51	0.53
3C–GaN	5.80	10.80	2.37
GaP	9.11	5.81	0.83
GaAs	10.90	4.97	0.52
GaSb	14.44	3.8	0.16
3C–InN	–	6.48	1.51
InP	9.61	5.04	0.86
InAs	12.25	4.20	0.50
InSb	15.68	3.33	0.22
2H–ZnO	3.72	12.94	3.04 ^b
ZnS	5.14	8.12	2.05
ZnSe	5.70	7.06	1.48
ZnTe	7.28	5.55	1.00
CdS	5.27	7.06	1.93
CdSe	6.10	6.16	1.53
CdTe	7.21	5.11	1.12
CuGaS ₂	6.15	7.46	1.43
CuInS ₂	6.3*	7.02	1.47
CuAlSe ₂	6.3*	6.85	1.25
CuGaSe ₂	7.3*	6.29	0.93
CuInSe ₂	9.00	5.34	0.75
CuGaTe ₂	8.0*	5.39	0.61
CuInTe ₂	9.20	4.78	0.55
AgGaSe ₂	6.80	5.96	1.09
AgInSe ₂	7.20	5.60	1.11
SiO ₂	2.10		3.99 ^c
Si ₃ N ₄	3.80		1.93 ^c
Al ₂ O ₃	3.13		3.23 ^c
ZrO ₂	4.84		≈ 3.2 ^c
HfO ₂	4.00		2.62 ^c

* $\epsilon_{\infty} = n^2$, ^a[8.24], ^b[8.25], ^c[8.26]

fit to the data of the zincblende-structure compound semiconductors [8.28]

$$\Phi_{\text{bp}}^{\text{p}} = 0.449 \cdot W_{\text{dg}} - [W_v(\Gamma) - W_v(k_{\text{mv}})]_{\text{ETB}}, \quad (8.14)$$

indicates that the branch points of these semiconductors lie 5% below the middle of the energy gap at the mean-value k -point. Table 8.1 displays the p-type branch-point energies of the Group IV elemental semiconductors, of SiC, and of III–V and II–VI compound semiconductors, as well as of some insulators.

A simple phenomenological model of Schottky contacts with a continuum of interface states and a constant density of states D_{is} across the semiconductor band gap yields the slope parameter [8.31, 32]

$$S_X = A_X / \left[1 + \left(e_0^2 / \varepsilon_i \varepsilon_0 \right) D_{is} \delta_{is} \right], \quad (8.15)$$

where ε_i is an interface dielectric constant. The parameter A_X depends on the electronegativity scale chosen and amounts to 0.86 eV/Miedema-unit and 1.79 eV/Pauling-unit. For $D_{is} \rightarrow 0$, relation (8.15) yields $S_X \rightarrow 1$ or, in other words, if no interface-induced gap states were present at the metal–semiconductor interfaces one would obtain the Schottky–Mott rule. The extension δ_{is} of the interface states may be approximated by their charge decay length $1/2q_{is}$. Mönch [8.32] used theoretical D_{gs}^{mi} and q_{gs}^{mi} data for metal-induced gap

states (MIGS), as the IFIGS in Schottky contacts are traditionally called, and plotted the $(e_0^2/\varepsilon_0)D_{gs}^{mi}/2q_{gs}^{mi}$ values versus the optical susceptibility $\varepsilon_\infty - 1$. The linear least-squares fit to the data points yielded [8.32]

$$A_X/S_X - 1 = 0.1 \cdot (\varepsilon_\infty - 1)^2, \quad (8.16)$$

where the reasonable assumption $\varepsilon_i \approx 3$ was made.

To a first approximation, the slope parameter D_X of heterostructure band offsets may be equated with the slope parameter S_X of Schottky contacts, since the IFIGS determine the intrinsic electric-dipole contributions to both the valence-band offsets and the barrier heights. Furthermore, the Group IV semiconductors and the elements constituting the III–V and II–VI compound semiconductors are all placed in the center columns of the Periodic Table and their electronegativities thus only differ by up to 10%. Consequently, the electric-dipole term $D_X \cdot (X_{sr} - X_{sl})$ may be neglected [8.9], so that (8.12) reduces to

$$\Delta W_v \cong \Phi_{bpr}^p - \Phi_{bpl}^p \quad (8.17)$$

for practical purposes.

8.3 Comparison of Experiment and Theory

8.3.1 Barrier Heights of Schottky Contacts

Experimental barrier heights of intimate, abrupt, clean and (above all) laterally homogeneous Schottky contacts on n-Si and n-GaAs as well as n-GaN, and the three SiC polytypes 3C, 6H and 4H are plotted in Figs. 8.10 and 8.11, respectively, versus the difference in the Miedema electronegativities of the metals and the semiconductors. Miedema's electronegativities [8.33, 34] are preferred since they were derived from properties of metal alloys and intermetallic compounds, while Pauling [8.8] considered covalent bonds in small molecules. The p- and n-type branch-point energies, $\Phi_{bp}^p = W_{bp} - W_v(\Gamma)$ and $\Phi_{bp}^n = W_c - W_{bp}$, respectively, add up to the fundamental band-gap energy $W_g = W_c - W_v(\Gamma)$. Hence, the barrier heights of n-type Schottky contacts are

$$\Phi_{Bn}^{hom} = \Phi_{bp}^n + S_X(X_m - X_s). \quad (8.18)$$

The electronegativity of a compound is taken as the geometric mean of the electronegativities of its constituent atoms.

First off all, the experimental data plotted in Figs. 8.10 and 8.11 clearly demonstrate that the different

experimental techniques, I/V , BEEM, IPEYS and PES, yield barrier heights of laterally homogeneous Schottky contacts which agree within the margins of experimental error.

Second, all experimental data are quantitatively explained by the branch-point energies (8.14) and the slope parameters (8.16) of the IFIGS-and-electronegativity theory. As was already mentioned in Sect. 8.1.1, the stacking fault, which is part of the interfacial Si(111)-(7×7)ⁱ reconstruction, causes an *extrinsic* electric dipole in addition to the *intrinsic* IFIGS electric dipole. The latter one is present irrespective of whether the interface structure is reconstructed or (1×1)ⁱ-unreconstructed. The extrinsic stacking fault-induced electric dipole quantitatively explains the experimentally observed barrier height lowering of 76 ± 2 meV.

Third, the IFIGS lines in Figs. 8.11a and 8.11b were drawn using the branch-point energies calculated for cubic 3C-GaN and 3C-SiC, respectively, since relation (8.12) was derived for zincblende-structure compounds only. However, the Schottky contacts were prepared on wurtzite-structure 2H-GaN and not just on cubic 3C-SiC but also on its hexagonal polytypes 4H and 6H. The good agreement between the experimen-

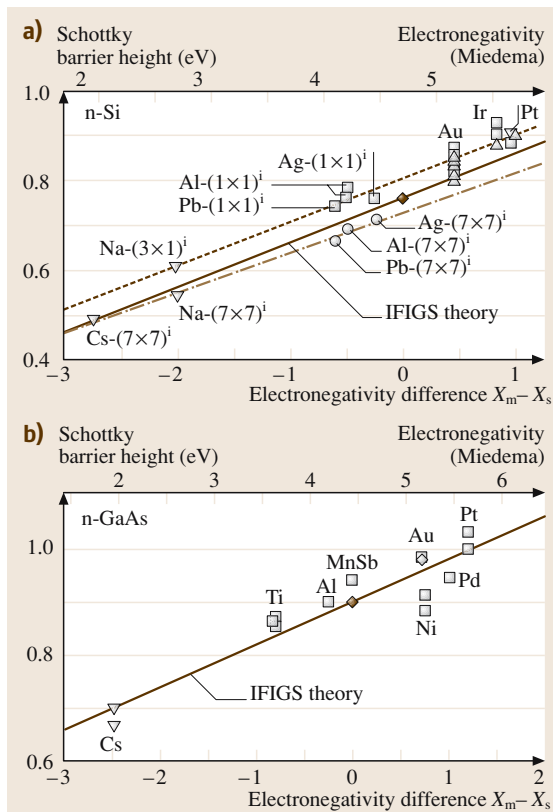


Fig. 8.10a,b Barrier heights of laterally homogeneous n-type silicon (a) and GaAs Schottky contacts (b) versus the difference in the Miedema electronegativities of the metals and the semiconductors. The \circ and \square , \diamond , Δ , and ∇ symbols differentiate the data from *I/V*, *BEEM*, *IPEYS*, and *PES* measurements, respectively. The *dashed* and the *dash-dotted* lines are the linear least-squares fits to the data from diodes with $(1 \times 1)^i$ -unreconstructed and $(7 \times 7)^i$ -reconstructed interfaces, respectively. The *solid IFIGS* lines are drawn with $S_X = 0.101$ eV/Miedema-unit and $\Phi_{bp}^p = 0.36$ eV for silicon (a) and with $S_X = 0.08$ eV/Miedema-unit and $\Phi_{bp}^p = 0.5$ eV for GaAs (b). After [8.14]

tal data and the *IFIGS* lines indicates that the p-type branch-point energies are rather insensitive to the specific bulk lattice structure of the semiconductor. This conclusion is further justified by the band-edge discontinuities of the semiconductor heterostructures, which were experimentally observed and are discussed in Sect. 8.3.2, and by the band-edge offsets of 3C/2H homostructures that were calculated for various semiconductors [8.35–39].

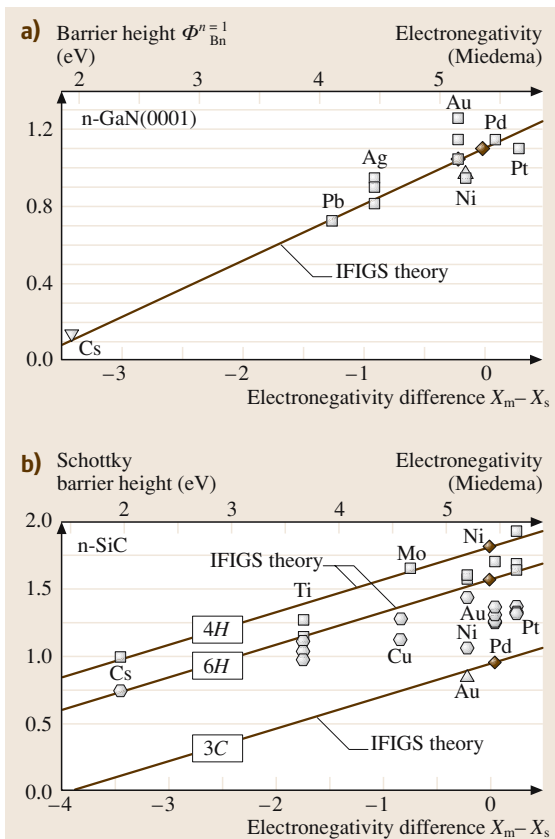


Fig. 8.11a,b Barrier heights of laterally homogeneous n-type GaN(0001) (a) and 3C-, 4H-, and 6H-SiC Schottky contacts (b) versus the difference in the Miedema electronegativities of the metals and the semiconductors. (a): The \square , \diamond , Δ , and ∇ symbols differentiate the data from *I/V*, *BEEM*, *IPEYS*, and *PES* measurements, respectively. The *solid IFIGS* line is drawn with $S_X = 0.29$ eV/Miedema-unit and $\Phi_{bp}^p = 2.37$ eV. (b): The \square , \diamond , and \circ symbols differentiate data of 4H-, 6H- and 3C-SiC Schottky contacts, respectively. The *solid IFIGS* lines are drawn with the band gaps of the polytypes minus $\Phi_{bp}^p = 1.44$ eV of cubic 3C-SiC and $S_X = 0.24$ eV/Miedema-unit. After [8.14]

8.3.2 Band Offsets of Semiconductor Heterostructures

In the bulk, and at interfaces of sp^3 -coordinated semiconductors, the chemical bonds are covalent. The simplest semiconductor–semiconductor interfaces are *lattice-matched* heterostructures. However, if the bond lengths of the two semiconductors differ then the interface will respond with tetragonal lattice distortions. Such

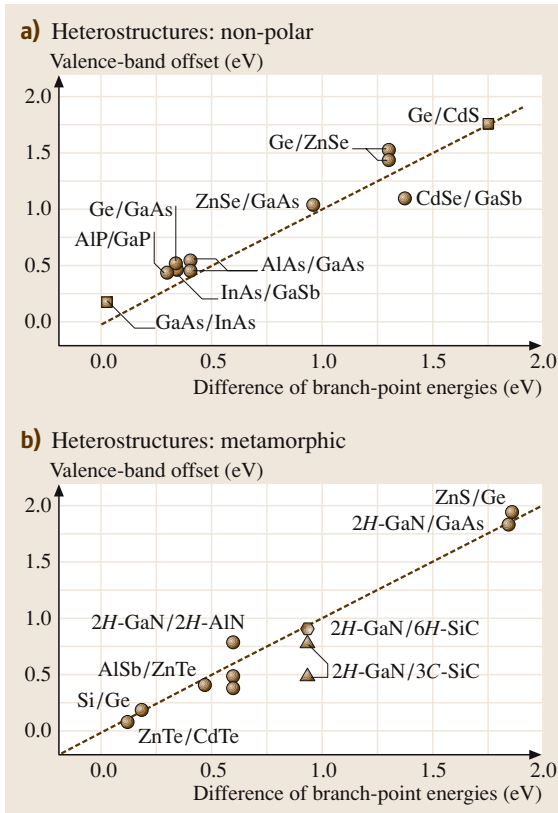


Fig. 8.12a,b Valence band offsets at nonpolar (110)-oriented (a) and metamorphic semiconductor heterostructures (b) versus the difference between the p-type branch-point energies of the semiconductors in contact. After [8.14]

pseudomorphic interfaces are under tensile or compressive stress. If the strain energy becomes too large then it is energetically more favorable to release the stress by the formation of misfit dislocations. Such *metamorphic* interfaces are almost relaxed.

In contrast to isovalent heterostructures, the chemical bonds at heterovalent interfaces require special attention, since interfacial donor- and acceptor-type bonds may cause interfacial electric dipoles [8.40]. No such extrinsic electric dipoles will exist normal to non-polar (110) interfaces. However, polar (001) interfaces behave quite differently. Acceptor bonds or donor bonds normal to the interface would exist at abrupt heterostructures. But, for reasons of charge neutrality, they have to be compensated by a corresponding density of donor bonds and acceptor bonds, respectively. This may be achieved by an intermixing at the interface which, on

the other hand, causes extrinsic electric dipoles. Their components normal to the interface will add an extrinsic electric-dipole contribution to the valence-band offset. In the following, only nonpolar, lattice-matched isovalent, and metamorphic heterostructures will be discussed.

The valence-band offsets at nonpolar, in other words (110)-oriented, heterostructures of compound semiconductors should equal the difference in the branch-point energies of the two semiconductors in contact provided the intrinsic **IFIGS** electric-dipole contribution can be neglected, see relation (8.17). Figure 8.12a displays respective experimental results for diamond- and zincblende-structure semiconductors as a function of the difference in the branch-point energies given in Table 8.1. The dashed line clearly demonstrates that the experimental data are excellently explained by the theoretical branch-point energies or, in other words, by the **IFIGS** theory.

As an example of lattice-matched and isovalent heterostructures, Fig. 8.13 shows valence-band offsets for $\text{Al}_{1-x}\text{Ga}_x\text{As}/\text{GaAs}$ heterostructures as a function of the alloy composition x . The **IFIGS** branch-point energies of the alloys were calculated assuming virtual $\text{Al}_{1-x}\text{Ga}_x$ cations [8.28], and were found to vary linearly as a function of composition between the values of AlAs and GaAs. More refined first-principles calculations yielded identical results [8.41, 42]. Figure 8.13 reveals that the theoretical **IFIGS** valence-band offsets fit the experimental data excellently.

Figure 8.12b displays valence-band offsets for metamorphic heterostructures versus the difference in the branch-point energies of the two semiconductors. The

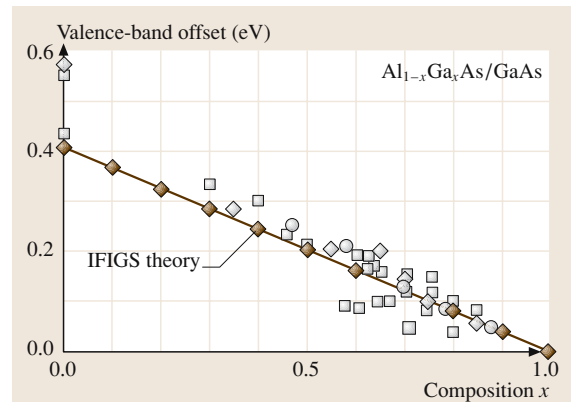


Fig. 8.13 Valence band offsets of lattice-matched and isovalent $\text{Al}_{1-x}\text{Ga}_x\text{As}/\text{GaAs}$ heterostructures as a function of alloy composition x . After [8.14]

dashed line indicates that the experimental results are again excellently described by the theoretical IFIGS data. This is true not only for heterostructures between cubic zincblende- and hexagonal wurtzite-structure compounds but also for wurtzite-structure Group III nitrides grown on both cubic 3C- and hexagonal 6H-SiC substrates. These observations suggest the following conclusions. First, all of the heterostructures considered in Fig. 8.12b are only slightly (if at all) strained, although their lattice parameters differ by up to 19.8%. Second, the calculations of the IFIGS branch-point energies assumed zincblende-structure semiconductors. These values, on the other hand, reproduce the experimental valence band offsets irrespective of whether the semiconductors have zincblende, wurtzite or, as in the case of 6H-SiC, another hexagonal-polytype structure. These findings again support the conclusion drawn from the GaN and SiC Schottky barrier heights in the previous section, that the IFIGS branch-point energies are rather insensitive to the specific semiconductor bulk lattice structure.

8.3.3 Band-Structure Lineup at Insulator Interfaces

The continuing miniaturization of complementary metal–oxide–semiconductor (CMOS) devices requires gate insulators where the dielectric constants (κ) are larger than the value of the silicon dioxide conventionally used. At present, the high- κ insulators Al_2O_3 , ZrO_2 , and HfO_2 are being intensively studied. Insulators may be considered to be wide-gap semiconductors. Hence, relations (8.11) and (8.12) also apply to insulator Schottky contacts and heterostructures. Unfortunately, the branch-point energies of these insulators cannot be obtained from relation (8.14) since it is valid for zincblende-structure compound semiconductors only. However, the experimental band offsets reported for SiO_2 , Si_3N_4 , Al_2O_3 , and HfO_2 heterostructures may be plotted as a function of the branch-point energies of the respective semiconductors [8.26]. Figure 8.14a reveals that the valence-band offsets become smaller with increasing branch-point energy of the semiconductors. Moreover, the data points reported for the many different SiO_2 heterostructures studied indicate a linear dependence for the valence-band offsets on the branch-point energy of the semiconductors, which may be written as

$$\Delta W_v = \varphi_{\text{vbo}} \left[\Phi_{\text{bp}}^{\text{p}}(\text{ins}) - \Phi_{\text{bp}}^{\text{p}}(\text{sem}) \right], \quad (8.19)$$

since the valence band offsets of insulator homostructures will definitely vanish. Such a linear relationship

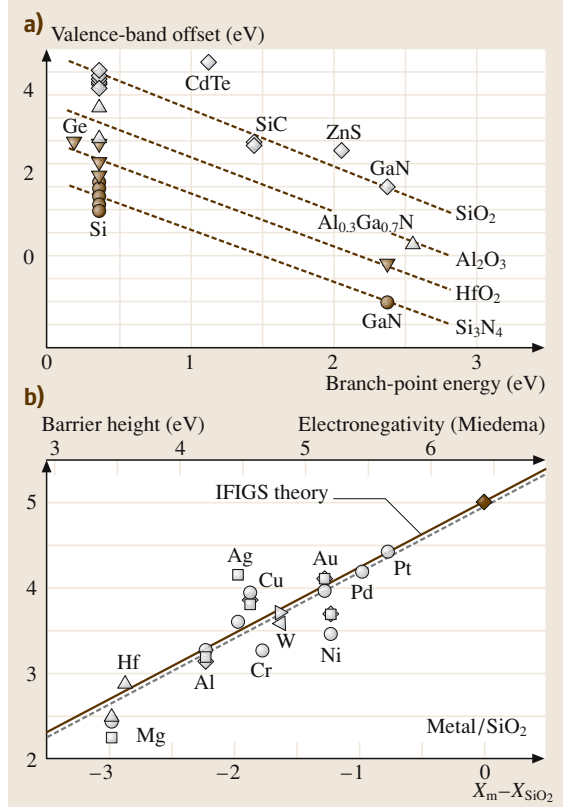


Fig. 8.14 (a) Valence band offsets of SiO_2 , Si_3N_4 , Al_2O_3 and HfO_2 heterostructures versus the p-type branch-point energies of the respective semiconductors. (b) n-type barrier heights of SiO_2 Schottky contacts versus the difference between electronegativities of the metal and SiO_2 . The dashed line is the linear least-squares fit to the data points. The solid IFIGS line is drawn with $\Phi_{\text{bp}}^{\text{p}} = 5 \text{ eV}$ ($W_g = 9 \text{ eV}$) and $S_X = 0.77 \text{ eV/Miedema-unit}$ ($\epsilon_\infty = 2.1$). After [8.25]

can also be adopted for the Al_2O_3 , HfO_2 and Si_3N_4 heterostructures, where less experimental results are available. Hence, the data displayed in Fig. 8.14a provide a means of determining the branch-point energies $\Phi_{\text{bp}}^{\text{p}}(\text{ins})$ of SiO_2 , Si_3N_4 , and the high- κ oxides Al_2O_3 and HfO_2 . The dashed lines in Fig. 8.14a are the linear least-squares fits to the respective data points. The experimental slope parameters φ_{vbo} range from 1.16 to 1.23 for HfO_2 and SiO_2 heterostructures, respectively, while relation (8.12) predicts $\varphi_{\text{vbo}} = 1$ provided that the electric dipole term $D_X \cdot (X_{\text{sr}} - X_{\text{sl}})$ vanishes. However, as well-established as this simplifying assumption is for the classical semiconductor heterostructures discussed in Sect. 8.3.2, it has ques-

tionable validity for the insulators considered here since they are much more ionic. Hence, the difference $\varphi_{\text{vbo}} - 1$ may be attributed to intrinsic electric-dipole layers at these insulator–semiconductor interfaces. The p-type branch-point energies $\Phi_{\text{bp}}^{\text{p}}$ of the insulators obtained from the linear least-squares fits are displayed in Table 8.1.

The reliability of these branch-point energies may be checked by, for example, analyzing barrier heights of respective insulator Schottky contacts. Such data are

only available for SiO_2 . Figure 8.14b displays the barrier heights of SiO_2 Schottky contacts as a function of the electronegativity difference $X_{\text{m}} - X_{\text{SiO}_2}$, where the electronegativity of SiO_2 is estimated as 6.42 Miedema-units. The linear least-squares fit

$$\Phi_{\text{Bn}} = (4.95 \pm 0.19) + (0.77 \pm 0.10) \times (X_{\text{m}} - X_{\text{SiO}_2})[\text{eV}] \quad (8.20)$$

to the experimental data agrees excellently with the prediction from the IFIGS-and-electronegativity theory.

8.4 Final Remarks

The local density approximation to density functional theory (LDA-DFT) is the most powerful and widely used tool in theoretical studies of the ground-state properties of solids. However, excitation energies such as the width of the energy gaps between the valence and conduction bands of semiconductors cannot be correctly obtained from such calculations. The fundamental band gaps of the elemental semiconductors C, Si and Ge as well as of the III–V and II–VI compounds are notoriously underestimated by 25 to 50%. However, it became possible to compute quasi-particle energies and band gaps of semiconductors from first principles using the so-called GW approximation for the electron self-energy [8.43, 44]. The resulting band gap energies agree to within 0.1 to 0.3 eV with experimental values.

For some specific metal–semiconductor contacts, the band-structure lineup was also studied by state-of-the-art ab-initio LDA-DFT calculations. The resulting LDA-DFT barrier heights were then subjected to a-posteriori corrections which consider quasi-particle effects and, if necessary, spin-orbit interactions and semicore-orbital effects. However, comparison of the theoretical results with experimental data gives an inconsistent picture. The *mean* values of the barrier heights of Al- and Zn/p-ZnSe contacts, which were calculated for different interface configurations using ab-initio LDA-DFT theory and a-posteriori spin-orbit and quasi-particle corrections [8.45, 46], agree with the experimental data to within the margins of experimental error. The same conclusion was reached for Al/ $\text{Al}_{1-x}\text{Ga}_x\text{As}$ Schottky

contacts [8.47]. However, ab-initio LDA-DFT barrier heights of Al-, Ag-, and Au/p-GaN contacts [8.48, 49], as well as of Al- and Ti/3C-SiC(001) interfaces [8.50, 51], strongly deviate from the experimental results.

As already mentioned, ab-initio LDF-DFT valence band offsets of $\text{Al}_{1-x}\text{Ga}_x\text{As}/\text{GaAs}$ heterostructures [8.41, 42] reproduce the experimental results well. The same holds for *mean* values of LDF-DFT valence-band offsets computed for different interface configurations of GaN- and AlN/SiC heterostructures [8.52–56].

The main difficulty which the otherwise extremely successful ab-initio LDF-DFT calculations encounter when describing semiconductor interfaces is not the precise exchange-correlation potential, which may be estimated in the GW approximation, but their remarkable sensitivity to the geometrical and compositional structure right at the interface. This aspect is more serious at metal–semiconductor interfaces than at heterostructures between two sp^3 -bonded semiconductors. The more conceptual IFIGS-and-electronegativity theory, on the other hand, quantitatively explains not only the barrier heights of ideal Schottky contacts but also the valence-band offsets of semiconductor heterostructures. Here again, the Schottky contacts are the more important case, since their zero-charge-transfer barrier heights equal the branch-point energies of the semiconductors, while the valence-band offsets are determined by the differences in the branch-point energies of the semiconductors in contact.

References

- | | | | |
|-----|--|-----|--|
| 8.1 | W. Schottky: <i>Naturwissenschaften</i> 26 , 843 (1938) | 8.4 | W. Schottky: <i>Phys. Zeitschr.</i> 41 , 570 (1940) |
| 8.2 | F. Braun: <i>Pogg. Ann. Physik Chemie</i> 153 , 556 (1874) | 8.5 | J. Bardeen: <i>Phys. Rev.</i> 71 , 717 (1947) |
| 8.3 | N. F. Mott: <i>Proc. Camb. Philos. Soc.</i> 34 , 568 (1938) | 8.6 | V. Heine: <i>Phys. Rev.</i> 138 , A1689 (1965) |

- 8.7 C. Tejedor, F. Flores: J. Phys. C **11**, L19 (1978)
- 8.8 L. N. Pauling: *The Nature of the Chemical Bond* (Cornell Univ. Press, Ithaca 1939)
- 8.9 W. Mönch: On the Present Understanding of Schottky Contacts. In: *Festkörperprobleme*, Vol. 26, ed. by P. Grosse (Vieweg, Braunschweig 1986) p. 67
- 8.10 W. Mönch: Phys. Rev. B **37**, 7129 (1988)
- 8.11 R. Schmitsdorf, T. U. Kampen, W. Mönch: Surf. Sci. **324**, 249 (1995)
- 8.12 J. L. Freeouf, T. N. Jackson, S. E. Laux, J. M. Woodall: Appl. Phys. Lett. **40**, 634 (1982)
- 8.13 W. Schottky: Phys. Zeitschr. **15**, 872 (1914)
- 8.14 W. Mönch: *Electronic Properties of Semiconductor Interfaces* (Springer, Berlin, Heidelberg 2004)
- 8.15 H.A. Bethe: MIT Radiation Lab. Rep. 43-12 (1942)
- 8.16 K. Takayanagi, Y. Tanishiro, M. Takahashi, S. Takahashi: Surf. Sci. **164**, 367 (1985)
- 8.17 H.-J. Im, Y. Ding, J. P. Pelz, W. J. Choyke: Phys. Rev. B **64**, 075310 (2001)
- 8.18 W. J. Kaiser, L. D. Bell: Phys. Rev. Lett. **60**, 1406 (1988)
- 8.19 L. D. Bell, W. J. Kaiser: Phys. Rev. Lett. **61**, 2368 (1988)
- 8.20 R. Turan, B. Aslan, O. Nur, M. Y. A. Yousif, M. Willander: Appl. Phys. A **72**, 587 (2001)
- 8.21 J. Cohen, J. Vilms, R. J. Archer: Hewlett-Packard R&D Report AFCRL-69-0287 (1969)
- 8.22 R. W. Grant, J. R. Waldrop, E. A. Kraut: Phys. Rev. Lett. **40**, 656 (1978)
- 8.23 S. B. Zhang, M. L. Cohen, S. G. Louie: Phys. Rev. B **34**, 768 (1986)
- 8.24 J. Tersoff: J. Vac. Sci. Technol. B **4**, 1066 (1986)
- 8.25 W. Mönch: Appl. Phys. Lett. **86**, 162101 (2005)
- 8.26 W. Mönch: Appl. Phys. Lett. **86**, 122101 (2005)
- 8.27 J. Tersoff: Phys. Rev. Lett. **52**, 465 (1984)
- 8.28 W. Mönch: J. Appl. Phys. **80**, 5076 (1996)
- 8.29 A. Baldereschi: Phys. Rev. B **7**, 5212 (1973)
- 8.30 D. R. Penn: Phys. Rev. **128**, 2093 (1962)
- 8.31 A. M. Cowley, S. M. Sze: J. Appl. Phys. **36**, 3212 (1965)
- 8.32 W. Mönch: Appl. Surf. Sci. **92**, 367 (1996)
- 8.33 A. R. Miedema, F. R. de Boer, P. F. de Châtel: J. Phys. F **3**, 1558 (1973)
- 8.34 A. R. Miedema, P. F. de Châtel, F. R. de Boer: Physica **100B**, 1 (1980)
- 8.35 A. Qteish, V. Heine, R. J. Needs: Phys. Rev. B **45**, 6534 (1992)
- 8.36 P. Käckell, B. Wenzien, F. Bechstedt: Phys. Rev. B **50**, 10761 (1994)
- 8.37 S. Ke, K. Zhang, X. Xie: J. Phys. Condens. Mat. **8**, 10209 (1996)
- 8.38 J. A. Majewski, P. Vogl: MRS Internet J. Nitride Semicond. Res. **3**, 21 (1998)
- 8.39 S.-H. Wei, S. B. Zhang: Phys. Rev. B **62**, 6944 (2000)
- 8.40 W. A. Harrison, E. A. Kraut, J. R. Waldrop, R. W. Grant: Phys. Rev. B **18**, 4402 (1978)
- 8.41 J. S. Nelson, A. F. Wright, C. Y. Fong: Phys. Rev. B **43**, 4908 (1991)
- 8.42 S. B. Zhang, M. L. Cohen, S. G. Louie, D. Tománek, M. S. Hybertsen: Phys. Rev. B **41**, 10058 (1990)
- 8.43 M. S. Hybertsen, S. G. Louie: Phys. Rev. B **34**, 5390 (1986)
- 8.44 R. W. Godby, M. Schlüter, L. J. Sham: Phys. Rev. B **37**, 10159 (1988)
- 8.45 M. Lazzarino, G. Scarel, S. Rubini, G. Bratina, L. Sorba, A. Franciosi, C. Berthod, N. Binggeli, A. Baldereschi: Phys. Rev. B **57**, R9431 (1998)
- 8.46 S. Rubini, E. Pellucchi, M. Lazzarino, D. Kumar, A. Franciosi, C. Berthod, N. Binggeli, A. Baldereschi: Phys. Rev. B **63**, 235307 (2001)
- 8.47 J. Bardi, N. Binggeli, A. Baldereschi: Phys. Rev. B **54**, R11102 (1996)
- 8.48 S. Picozzi, A. Continenza, G. Satta, S. Massidda, A. J. Freeman: Phys. Rev. B **61**, 16736 (2000)
- 8.49 S. Picozzi, G. Profeta, A. Continenza, S. Massidda, A. J. Freeman: Phys. Rev. B **65**, 165316 (2002)
- 8.50 J. Hoekstra, M. Kohyama: Phys. Rev. B **57**, 2334 (1998)
- 8.51 M. Kohyama, J. Hoekstra: Phys. Rev. B **61**, 2672 (2000)
- 8.52 M. Städele, A. J. Majewski, P. Vogl: Phys. Rev. B **56**, 6911 (1997)
- 8.53 J. A. Majewski, M. Städele, P. Vogl: Mater. Res. Soc. Symp. Proc. **449**, 917 (1997)
- 8.54 N. Binggeli, P. Ferrara, A. Baldereschi: Phys. Rev. B **63**, 245306 (2001)
- 8.55 B. K. Agrawal, S. Agrawal, R. Srivastava, P. Srivastava: Physica E **11**, 27 (2001)
- 8.56 M. R. Laridjani, P. Masri, J. A. Majewski: Mater. Res. Soc. Symp. Proc. **639**, G11.34 (2001)

9. Charge Transport in Disordered Materials

This chapter surveys general theoretical concepts developed to qualitatively understand and to quantitatively describe the electrical conduction properties of disordered organic and inorganic materials. In particular, these concepts are applied to describe charge transport in amorphous and microcrystalline semiconductors and in conjugated and molecularly doped polymers. Electrical conduction in such systems is achieved through incoherent transitions of charge carriers between spatially localized states. Basic theoretical ideas developed to describe this type of electrical conduction are considered in detail. Particular attention is given to the way the kinetic coefficients depend on temperature, the concentration of localized states, the strength of the applied electric field, and the charge carrier localization length. Charge transport via delocalized states in disordered systems and the relationships between kinetic coefficients under the nonequilibrium conditions are also briefly reviewed.

9.1	General Remarks on Charge Transport in Disordered Materials	163
9.2	Charge Transport in Disordered Materials via Extended States	167
9.3	Hopping Charge Transport in Disordered Materials via Localized States	169
9.3.1	Nearest-Neighbor Hopping.....	170
9.3.2	Variable-Range Hopping.....	172
9.3.3	Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Inorganic Noncrystalline Materials.....	173
9.3.4	Description of Charge Carrier Energy Relaxation and Hopping Conduction in Organic Noncrystalline Materials.....	180
9.4	Concluding Remarks	184
	References	185

Many characteristics of charge transport in disordered materials differ markedly from those in perfect crystalline systems. The term “disordered materials” usually refers to noncrystalline solid materials without perfect order in the spatial arrangement of atoms. One should distinguish between disordered materials with ionic conduction and those with electronic conduction. Disordered materials with ionic conduction include various glasses consisting of a “network-formers” such as SiO_2 , B_2O_3 and Al_2O_3 , and of “network-modifiers” such as Na_2O , K_2O and Li_2O . When an external voltage is applied, ions can drift by hopping over potential barriers in the glass matrix, contributing to the electrical conduction of the material. Several fascinating effects have been observed for this kind of electrical conduction. One is the extremely nonlinear dependence of the conductivity on the concentration of ions in the material. Another beautiful phenomenon is the so-called “mixed alkali effect”: mixing two different modifiers in one glass leads to an enormous drop in the conductivity in comparison to that

of a single modifier with the same total concentration of ions. A comprehensive description of these effects can be found in the review article of Bunde et al. [9.1]. Although these effects sometimes appear puzzling, they can be naturally and rather trivially explained using routine classical percolation theory [9.2]. The description of ionic conduction in glasses is much simplified by the inability of ions to tunnel over large distances in the glass matrix in single transitions. Every transition occurs over a rather small interatomic distance, and it is relatively easy to describe such electrical conductivity theoretically [9.2]. On the other hand, disordered systems with electronic conduction have a much more complicated theoretical description. Transition probabilities of electrons between spatially different regions in the material significantly depend not only on the energy parameters (as in the case of ions), but also on spatial factors such as the tunnelling distance, which can be rather large. The interplay between the energy and spatial factors in the transition probabilities of electrons makes the develop-

ment of a theory of electronic conduction in disordered systems challenging. Since the description of electronic conduction is less clear than that of ionic conduction, and since disordered electronic materials are widely used for various device applications, in this chapter we concentrate on disordered materials with the electronic type of electrical conduction.

Semiconductor glasses form one class of such materials. This class includes amorphous selenium, a-Se and other chalcogenide glasses, such as a-As₂Se₃. These materials are usually obtained by quenching from the melt. Another broad class of disordered materials, inorganic amorphous semiconductors, includes amorphous silicon a-Si, amorphous germanium a-Ge, and their alloys. These materials are usually prepared as thin films by the deposition of atomic or molecular species. Hydrogenated amorphous silicon, a-Si:H, has attracted much attention from researchers, since incorporation of hydrogen significantly improves conduction, making it favorable for use in amorphous semiconductor devices. Many other disordered materials, such as hydrogenated amorphous carbon (a-C:H) and its alloys, polycrystalline and microcrystalline silicon are similar to a-Si:H in terms of their charge transport properties. Some crystalline materials can also be considered to be disordered systems. This is the case for doped crystals if transport phenomena within them are determined by randomly distributed impurities, and for mixed crystals with disordered arrangements of various types of atoms in the crystalline lattice. In recent years much research has also been devoted to the study of organic disordered materials, such as conjugated and molecularly doped polymers and organic glasses, since these systems has been shown to possess electronic properties similar to those of inorganic disordered materials, while they are easier to manufacture than the latter systems.

There are two reasons for the great interest of researchers in the conducting properties of disordered materials. On the one hand, disordered systems represent a challenging field in a purely academic sense. For many years the theory of how semiconductors perform charge transport was mostly confined to crystalline systems where the constituent atoms are in regular arrays. The discovery of how to make solid amorphous materials and alloys led to an explosion in measurements of the electronic properties of these new materials. However, the concepts often used in textbooks to describe charge carrier transport in crystalline semiconductors are based on an assumption of long-range order, and so they cannot be applied to electronic transport in disor-

dered materials. It was (and still is) a highly challenging task to develop a consistent theory of charge transport in such systems. On the other hand, the explosion in research into charge transport in disordered materials is related to the various current and potential device applications of such systems. These include the application of disordered inorganic and organic materials in photovoltaics (the functioning material in solar cells), in electrophotography, in large-area displays (they are used in thin film transistors), in electrical switching threshold and memory devices, in light-emitting diodes, in linear image sensors, and in optical recording devices. Readers interested in the device applications of disordered materials should be aware that there are numerous monographs on this topic: the literature on this field is very rich. Several books are recommended (see [9.3–12]), as are numerous review articles referred to in these books.

In this chapter we focus on disordered semiconductor materials, ignoring the broad class of disordered metals. In order to describe electronic transport in disordered metals, one can more or less successfully apply extended and modified conventional theoretical concepts developed for electron transport in ordered crystalline materials, such as the Boltzmann kinetic equation. Therefore, we do not describe electronic transport in disordered metals here. We can recommend a comprehensive monograph to interested readers (see [9.13]), in which modern concepts about conduction in disordered metals are presented beautifully.

Several nice monographs on charge transport in disordered semiconductors are also available. Although many of them were published several years ago (some even decades ago), we can recommend them to the interested reader as a source of information on important experimental results. These results have permitted researchers the present level of understanding of transport phenomena in disordered inorganic and organic materials. A comprehensive collection of experimental data for noncrystalline materials from the books specified above would allow one to obtain a picture of the modern state of experimental research in the field.

We will focus in this chapter on the theoretical description of charge transport in disordered materials, introducing some basic concepts developed to describe electrical conduction. Several excellent books already exist in which a theoretical description of charge transport in disordered materials is the main topic. Among others we can recommend the books of *Shklovskii* and *Efros* [9.14], *Zvyagin* [9.15], *Böttger* and *Bryksin* [9.16], and *Overhof* and *Thomas* [9.17]. There appears to be

a time gap in which comprehensive monographs on the theoretical description of electrical conduction in disordered materials were not published. During this period

some new and rather powerful theoretical concepts were developed. We present these concepts below, along with some more traditional ones.

9.1 General Remarks on Charge Transport in Disordered Materials

Although the literature on transport phenomena in disordered materials is enormously rich, there are still many open questions in this field due to various problems specific to such materials. In contrast to ordered crystalline semiconductors with well-defined electronic energy structures consisting of energy bands and energy gaps, the electronic energy spectra of disordered materials can be treated as quasi-continuous. Instead of bands and gaps, one can distinguish between extended and localized states in disordered materials. In an extended state, the charge carrier wavefunction is spread over the whole volume of a sample, while the wavefunction of a charge carrier is localized in a spatially restricted region in a localized state, and a charge carrier present in such a state cannot spread out in a plane wave as in ordered materials. Actually, localized electron states are known in ordered systems too. Electrons and holes can be spatially localized when they occupy donors or acceptors or some other impurity states or structural defects in ordered crystalline materials. However, the localized states usually appear as δ -like discrete energy levels in the energy spectra of such materials. In disordered semiconductors, on the other hand, energy levels related to spatially localized states usually fill the energy spectrum continuously. The energy that separates the extended states from the localized ones in disordered materials is called the mobility edge. To be precise, we will mostly consider the energy states for electrons in the following. In this case, the states above the mobility edge are extended and the states below the edge are localized. The localized states lie energetically above the extended states for holes. The energy region between the mobility edges for holes and electrons is called the mobility gap. The latter is analogous to the band gap in ordered systems, although the mobility gap contains energy states, namely the spatially localized states. Since the density of states (DOS), defined as the number of states per unit energy per unit volume, usually decreases when the energy moves from the mobility edges toward the center of the mobility gap, the energy regions of localized states in the vicinity of the mobility edges are called band tails. We would like to emphasize that the charge transport properties depend significantly

on the energy spectrum in the vicinity and below the mobility edge (in the band tails). Unfortunately this energy spectrum is not known for almost all disordered materials. A whole variety of optical and electrical investigation techniques have proven unable to determine this spectrum. Since the experimental information on this spectrum is rather vague, it is difficult to develop a consistent theoretical description for charge transport *ab initio*. The absence of reliable information on the energy spectrum and on the structures of the wavefunctions in the vicinity and below the mobility edges can be considered to be the main problem for researchers attempting to quantitatively describe the charge transport properties of disordered materials.

An overview of the energy spectrum in a disordered semiconductor is shown in Fig. 9.1. The energy levels ε_v and ε_c denote the mobility edges for the valence and conduction bands, respectively. Electron states in the mobility gap between these energies are spatially localized. The states below ε_v and above ε_c can be occupied by delocalized holes and electrons. Some peaks in the DOS are shown in the mobility gap, which can be created by some defects with particularly high concentrations. Although there is a consensus between researchers on the general view of the DOS in disordered materials,

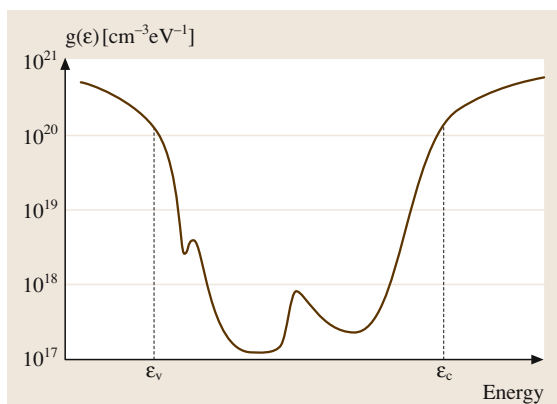


Fig. 9.1 Density of states of a noncrystalline semiconductor (schematic); ε_v and ε_c correspond to mobility edges in the conduction band and the valence band, respectively

the particular structure of the energy spectrum is not known for most disordered systems. From a theoretical point of view, it is enormously difficult to calculate this spectrum.

There are several additional problems that make the study of charge transport in disordered materials more difficult than in ordered crystalline semiconductors. The particular spatial arrangements of atoms and molecules in different samples with the same chemical composition can differ from each other depending on the preparation conditions. Hence, when discussing electrical conduction in disordered materials one often should specify the preparation conditions. Another problem is related to the long-time relaxation processes in disordered systems. Usually these systems are not in thermodynamic equilibrium and the slow relaxation of the atoms toward the equilibrium arrangement can lead to some changes in electrical conduction properties. In some disordered materials a long-time electronic relaxation can affect the charge transport properties too, particularly at low temperatures, when electronic spatial rearrangements can be very slow. At low temperatures, when tunneling electron transitions between localized states dominate electrical conduction, this long-time electron relaxation can significantly affect the charge transport properties.

It is fortunate that, despite these problems, some general transport properties of disordered semiconductors have been established. Particular attention is usually paid to the temperature dependence of the electrical conductivity, since this dependence can indicate the underlying transport mechanism. Over a broad temperature range, the direct current (DC) conductivity in disordered materials takes the form

$$\sigma = \sigma_0 \exp \left[- \left(\frac{\Delta(T)}{k_B T} \right)^\beta \right], \quad (9.1)$$

where the pre-exponential factor σ_0 depends on the underlying system and the power exponent β depends on the material and also sometimes on the temperature range over which the conductivity is studied; $\Delta(T)$ is the activation energy. In many disordered materials, like vitreous and amorphous semiconductors, σ_0 is of the order of 10^2 – $10^4 \Omega^{-1} \text{cm}^{-1}$. In such materials the power exponent β is close to unity at temperatures close to and higher than the room temperature, while at lower temperatures β can be significantly smaller than unity. In organic disordered materials, values of β that are larger than unity also have been reported. For such systems the value $\beta \approx 2$ is usually considered to be appropriate [9.18].

Another important characteristic of the electrical properties of a disordered material is its alternating current (AC) conductivity measured when an external alternating electric field with some frequency ω is applied. It has been established in numerous experimental studies that the real part of the AC conductivity in most disordered semiconductors depends on the frequency according to the power law

$$\text{Re } \sigma(\omega) = C \omega^s, \quad (9.2)$$

where C is constant and the power s is usually smaller than unity. This power law has been observed in numerous materials at different temperatures over a wide frequency range. This frequency dependence differs drastically from that predicted by the standard kinetic theory developed for quasi-free charge carriers in crystalline systems. In the latter case, the real part of the AC conductivity has the frequency dependence

$$\text{Re } \sigma(\omega) = \frac{ne^2}{m} \frac{\tau}{1 + \omega^2 \tau^2}, \quad (9.3)$$

where n is the concentration of charge carriers, e is the elementary charge, m is the effective mass and τ is the momentum relaxation time. Since the band electrons in crystalline semiconductors usually have rather short momentum relaxation times, $\tau \approx 10^{-14}$ s, the contribution of charge carriers in delocalized states to the AC conductivity usually does not depend on frequency at $\omega \ll \tau^{-1}$. Therefore, the observed frequency dependence described by (9.2) should be ascribed to the contribution of charge carriers in localized states.

One of the most powerful tools used to study the concentrations of charge carriers and their mobilities in crystalline semiconductors is the provided by measurements of the Hall constant, R_H . Such measurements also provide direct and reliable information about the sign of the charge carriers in crystalline materials. Unfortunately, this is not the case for disordered materials. Moreover, several anomalies have been established for Hall measurements in the latter systems. For example, the sign of the Hall constant in disordered materials sometimes differs from that of the thermoelectric power, α . This anomaly has not been observed in crystalline materials. The anomaly has been observed in liquid and solid noncrystalline semiconductors. Also, in some materials, like amorphous arsenic, a-As, $R_H > 0$, $\alpha < 0$, while in many other materials other combinations with different signs of R_H and α have been experimentally established.

In order to develop a theoretical picture of the transport properties of any material, the first issues to clarify

are the spectrum of the energy states for charge carriers and the spatial structure of such states. Since these two central issues are yet to be answered properly for noncrystalline materials, the theory of charge transport in disordered systems should be considered to be still in its embryonic stage.

The problem of deducing electron properties in a random field is very complicated, and the solutions obtained so far only apply to some very simple models. One of them is the famous Anderson model that illustrates the localization phenomenon caused by random disorder [9.19]. In this model, one considers a regular system of rectangular potential wells with randomly varying depths, as shown schematically in Fig. 9.2. The ground state energies of the wells are assumed to be randomly distributed over the range with a width of W . First, one considers the ordered version of the model, with W equal to zero. According to conventional band theory, a narrow band arises in the ordered system where the energy width depends on the overlap integral I between the electron wavefunctions in the adjoining

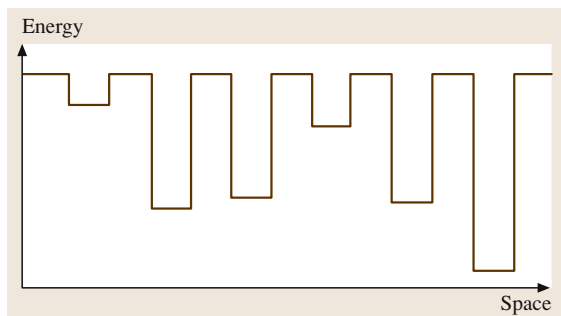


Fig. 9.2 Anderson model of disorder potential

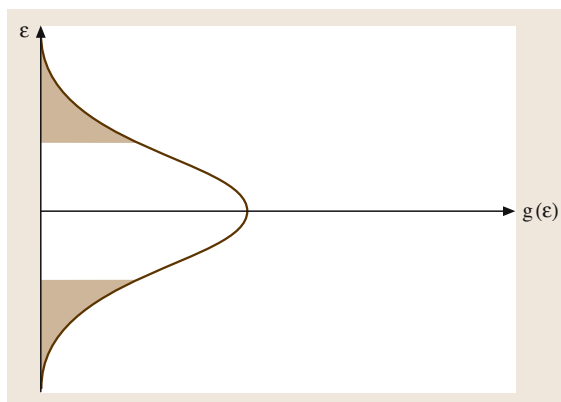


Fig. 9.3 Density of states in the Anderson model. *Hatched regions in the tails correspond to spatially localized states*

wells. The eigenstates in such a model are delocalized with wavefunctions of the Bloch type. This is trivial. The problem is to find the solution for a finite degree of disorder ($W \neq 0$). The result from the Anderson model for such a case is described as follows. At some particular value for the ratio $W/(zI)$, where z is the coordination number of the lattice, all electron states of the system are spatially localized. At smaller values of $W/(zI)$ some states in the outer regions of the DOS are localized and other states in the middle of the DOS energy distribution are spatially extended, as shown schematically in Fig. 9.3. This is one of the most famous results in the transport theory of disordered systems. When considering this result, one should note the following points. (i) It was obtained using a single-electron picture without taking into account long-range many-particle interactions. However, in disordered systems with localized electrons such interactions can lead to the localization of charge carriers and they often drastically influence the energy spectrum [9.14]. Therefore the applicability of the single-electron Anderson result to real systems is questionable. (ii) Furthermore, the energy structure of the Anderson model shown in Fig. 9.3 strongly contradicts that observed in real disordered materials. In real systems, the mobility gap is located between the mobility edges, as shown in Fig. 9.1, while in the Anderson model the energy region between the mobility edges is filled with delocalized states. Moreover, in one-dimensional and in some two-dimensional systems, the Anderson model predicts that all states are localized at any amount of disorder. These results are of little help when attempting to interpret the DOS scheme in Fig. 9.1.

A different approach to the localization problem is to try to impose a random potential $V(x)$ onto the band structure obtained in the frame of a traditional band theory. Assuming a classical smoothly varying (in space)



Fig. 9.4a–c Disorder potential landscape experienced by a charge carrier (a). Regions with energies below some given energy level E_c are colored black. In frame (b) this level is very low and there is no connected path through the system via black regions. In frame (c) the level E_c corresponds to the classical percolation level

potential $V(x)$ with a Gaussian distribution function

$$F(V) = \frac{1}{\varepsilon_0 \sqrt{2\pi}} \exp\left(-\frac{V^2}{2\varepsilon_0^2}\right), \quad (9.4)$$

one can solve the localization problem using the classical percolation theory illustrated in Fig. 9.4. In Fig. 9.4a, an example of a disorder potential experienced by electrons is shown schematically. In Fig. 9.4b and Fig. 9.4c the regions below a given energy level E_c are colored black. In Fig. 9.4b this level is positioned very low, so that regions with energies below E_c do not provide a connected path through the system. In Fig. 9.4c an infinite percolation cluster consisting only of black regions exists. The E_c that corresponds to the first appearance of such a connected path is called the classical percolation level [9.14]. Mathematically solving the percolation problem shows that the mobility edge identified with the classical percolation level in the potential $V(x)$ is shifted with respect to the band edge of the ordered system by an amount $\xi\varepsilon_0$, where $\xi \approx 0.96$ towards the center of the bandgap [9.15]. A similar result, though with a different constant ξ , can be obtained via a quantum-mechanical treatment of a short-range potential $V(x)$ of white-noise type [9.20]. As the amplitude ε_0 of the random potential increases the band gap narrows, while the conduction and valence bands become broader. Although this result is provided by both limiting models – by the classical one with a long-range smoothly varying potential $V(x)$ and by the quantum-mechanical one with a short-range white-noise potential $V(x)$ – none of the existing theories can reliably describe the energy spectrum of a disordered material and the properties of the charge carrier wavefunctions in the vicinity of the mobility edges, in other words in the energy range which is most important for charge transport.

The DC conductivity can generally be represented in the form

$$\sigma = e \int \mu(\varepsilon) n(\varepsilon) d\varepsilon, \quad (9.5)$$

where e is the elementary charge, $n(\varepsilon)d\varepsilon$ is the concentration of electrons in the energy range between ε and $\varepsilon + d\varepsilon$ and $\mu(\varepsilon)$ is the mobility of these electrons. The integration is carried out over all energies ε . Under equilibrium conditions, the concentration of electrons $n(\varepsilon)d\varepsilon$ is determined by the density of states $g(\varepsilon)$ and the Fermi function $f(\varepsilon)$, which depends on the position of the Fermi energy ε_F (or a quasi-Fermi energy in the case of the stationary excitation of electrons):

$$n(\varepsilon) = g(\varepsilon) f(\varepsilon), \quad (9.6)$$

where

$$f(\varepsilon) = \frac{1}{1 + \exp\left(\frac{\varepsilon - \varepsilon_F}{k_B T}\right)}. \quad (9.7)$$

Here T is the temperature and k_B is the Boltzmann constant.

The Fermi level in almost all known disordered semiconductors under real conditions is situated in the mobility gap – in the energy range which corresponds to spatially localized electron states. The charge carrier mobility $\mu(\varepsilon)$ in the localized states below the mobility edge is much less than that in the extended states above the mobility edge. Therefore, at high temperatures, when a considerable fraction of electrons can be found in the delocalized states above the mobility edge, these states dominate the electrical conductivity of the system. The corresponding transport mechanism under such conditions is similar to that in ordered crystalline semiconductors. Electrons in the states within the energy range of the width, of the order $k_B T$ above the mobility edge, dominate the conductivity. In such a case the conductivity can be estimated as

$$\sigma \approx e \mu_c n(\varepsilon_c) k_B T, \quad (9.8)$$

where μ_c is the electron mobility in the states above the mobility edge ε_c , and $n(\varepsilon_c)k_B T$ is their concentration. This equation is valid under the assumption that the typical energy scale of the DOS function $g(\varepsilon)$ above the mobility edge is larger than $k_B T$. The position of the Fermi level in disordered materials usually depends on temperature only slightly. Combining (9.6)–(9.8), one obtains the temperature dependence of the DC conductivity in the form

$$\sigma = \sigma_0 \exp\left(-\frac{\Delta}{k_B T}\right), \quad (9.9)$$

described by (9.1) with $\beta = 1$ and constant activation energy, which is observed in most disordered semiconductors at high temperatures.

In order to obtain the numerical value of the conductivity in this high-temperature regime, one needs to know the density of states in the vicinity of the mobility edge $g(\varepsilon_c)$, and also the magnitude of the electron mobility μ_c in the delocalized states above ε_c . While the magnitude of $g(\varepsilon_c)$ is usually believed to be close to the DOS value in the vicinity of the band edge in crystalline semiconductors, there is no consensus among researchers on the magnitude of μ_c . In amorphous semiconductors μ_c is usually estimated to be in the range of $1 \text{ cm}^2/\text{Vs}$ to $10 \text{ cm}^2/\text{Vs}$. Unfortunately, there are no reliable theoretical calculations of this quantity for most disordered

materials. The only exception is provided by so-called mixed crystals, which are also sometimes called crystalline solid solutions. In the next section we describe the theoretical method which allows one to estimate μ_c in

such systems. This method can be extended to other disordered materials, provided the statistical properties of the disorder potential, essential for electron scattering, are known.

9.2 Charge Transport in Disordered Materials via Extended States

The states with energies below ε_v and above ε_c in disordered materials are believed to possess similar properties to those of extended states in crystals. Various experimental data suggest that these states in disordered materials are delocalized states. However, traditional band theory is largely dependent upon the system having translational symmetry. It is the periodic atomic structure of crystals that allows one to describe electrons and holes within such a theory as quasi-particles that exhibit behavior similar to that of free particles in vacuum, albeit with a renormalized mass (the so-called “effective mass”). The energy states of such quasi-particles can be described by their momentum values. The wavefunctions of electrons in these states (the so-called Bloch functions) are delocalized. This means that the probability of finding an electron with a given momentum is equal at corresponding points of all elementary cells of the crystal, independent on the distance between the cells.

Strictly speaking, the traditional band theory fails in the absence of translational symmetry – for disordered systems. Nevertheless, one still assumes that the charge carriers present in delocalized states in disordered materials can be approximately described by wavefunctions with a spatially homogeneous probability of finding a charge carrier with a given quasi-momentum. As for crystals, one starts from the quasi-free particle picture and considers the scattering effects in a perturbation approach following the Boltzmann kinetic description. This description is valid if the de Broglie wavelength of the charge carrier $\lambda = \hbar/p$ is much less than the mean free path $l = v\tau$, where τ is the momentum relaxation time and p and v are the characteristic values of the momentum and velocity, respectively. This validity condition for the description based on the kinetic Boltzmann equation can also be expressed as $\hbar/\tau \ll \varepsilon$, where ε is the characteristic kinetic energy of the charge carriers, which is equal to $k_B T$ for a nondegenerate electron gas and to the Fermi energy in the degenerate case. While this description seems valid for delocalized states far from the mobility edges, it fails for energy states in the vicinity of the mobility edges. So far, there has been

no consensus between the theorists on how to describe charge carrier transport in the latter states. Moreover, it is not clear whether the energy at which the carrier mobility drops coincides with the mobility edge or whether it is located above the edge in the extended states. Numerous discussions of this question, mostly based on the scaling theory of localization, can be found in special review papers. For the rest of this section, we skip this rather complicated subject and instead we focus on the description of charge carrier transport in a semiconductor with a short-range random disorder potential of white-noise type. This seems to be the only disordered system where a reliable theory exists for charge carrier mobility via extended states above the mobility edge. Semiconductor solid solutions provide an example of a system with this kind of random disorder [9.20–25].

Semiconductor solid solutions $A_x B_{1-x}$ (mixed crystals) are crystalline semiconductors in which the sites of the crystalline sublattice can be occupied by atoms of two different types, A and B. Each site can be occupied by either an A or a B atom with some given probability x between zero and unity. The value x is often called the composition of the material. Due to the random spatial distributions of the A and B atoms, local statistical fluctuations in the composition inside the sample are unavoidable, meaning that mixed crystals are disordered systems. Since the position of the band edge depends on the composition x , these fluctuations in local x values lead to the disorder potential for electrons and holes within the crystal. To be precise, we will consider the influence of the random potential on a conduction band electron. Let $E_c(x)$ be the conduction band minimum for a crystal with composition x . In Fig. 9.5 a possible schematic dependence $E_c(x)$ is shown. If the average composition for the whole sample is x_0 , the local positions of the band edge $E_c(x)$ fluctuate around the average value $E_c(x_0)$ according to the fluctuations of the composition x around x_0 . For small deviations in composition Δx from the average value, one can use the linear relation

$$E_c(x_0 + \Delta x) = E_c(x_0) + \alpha \Delta x, \quad (9.10)$$

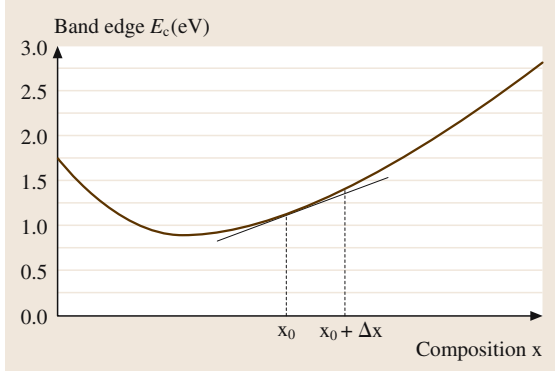


Fig. 9.5 Schematic dependence of the conduction band edge ε_c on composition x in a mixed crystal A_xB_{1-x}

where

$$\alpha = \left. \frac{dE_c(x)}{dx} \right|_{x=x_0}. \quad (9.11)$$

If the deviation of the concentration of A atoms from its mean value in some region of a sample is $\xi(r)$ and the total concentration of (sub)lattice sites is N , the deviation of the composition in this region is $\Delta x = \xi(r)/N$, and the potential energy of an electron at the bottom of the conduction band is

$$V(r) = \alpha \frac{\xi(r)}{N}. \quad (9.12)$$

Although one calls the disorder in such systems a “short-range” disorder, it should be noted that the consideration is valid only for fluctuations that are much larger than the lattice constant of the material. The term “short-range” is due to the assumption that the statistical properties of the disorder are absolutely uncorrelated. This means that potential amplitudes in the adjoining spatial points are completely uncorrelated to each other. Indeed, it is usually assumed that the correlation function of the disorder in mixed crystals can be approximated by a white-noise correlation function of the form

$$\langle \xi(r)\xi(r') \rangle = x(1-x)N\delta(r-r'). \quad (9.13)$$

The random potential caused by such compositional fluctuations is then described by the correlation function [9.20]

$$\langle V(r)V(r') \rangle = \gamma\delta(r-r') \quad (9.14)$$

with

$$\gamma = \frac{\alpha^2}{N}x(1-x). \quad (9.15)$$

Charge carriers in mixed crystals are scattered by compositional fluctuations. As is usual in kinetic descriptions of free electrons, the fluctuations on the spatial scale of the order of the electron wavelength are most efficient. Following *Shlimak et al.* [9.23], consider an isotropic quadratic energy spectrum

$$\varepsilon_p = \frac{p^2}{2m}, \quad (9.16)$$

where p and m are the quasi-momentum and the effective mass of an electron, respectively. The scattering rate for such an electron is

$$v_p = \frac{2\pi}{\hbar} \sum_q \langle |V_q|^2 \rangle (1 - \cos \vartheta_q) \delta(\varepsilon_p - \varepsilon_{p-q}), \quad (9.17)$$

where ϑ_q is the scattering angle and

$$\langle |V_q|^2 \rangle = \frac{1}{\Omega} \int d^3r \exp(iqr) \langle V(r)V(0) \rangle. \quad (9.18)$$

The quantity Ω in this formula is the normalization volume. Using the correlation function (9.14), one obtains the relation

$$\langle |V_q|^2 \rangle = \frac{\alpha^2 x(1-x)}{\Omega N}, \quad (9.19)$$

which shows that the scattering by compositional fluctuations is equivalent to that by a short-range potential [9.23]. Substituting (9.19) into (9.17) one obtains the following expression for the scattering rate [9.20]

$$v_p = \frac{\alpha^2 x(1-x)mp}{\pi \hbar^4 N}. \quad (9.20)$$

This formula leads to an electron mobility of the following form in the framework of the standard Drude approach [9.20, 23]

$$\mu_C = \frac{\pi^{3/2}}{2\sqrt{2}} \frac{e\hbar^4 N}{\alpha^2 x(1-x)m^{5/2}(k_B T)^{1/2}}. \quad (9.21)$$

Very similar formulae can be found in many recent publications (see for example *Fahy and O'Reilly* [9.26]). It has also been modified and applied to two-dimensional systems [9.27] and to disordered diluted magnetic semiconductors [9.28].

It would not be difficult to apply this theoretical description to other disordered systems, provided the correlation function of the disorder potential takes the form of (9.14) with known amplitude γ . However, it is worth emphasizing that the short-range disorder of white-noise type considered here is a rather simple

model that cannot be applied to most disordered materials. Therefore, we can conclude that the problem of theoretically describing charge carrier mobility via delocalized states in disordered materials is still waiting to be solved.

9.3 Hopping Charge Transport in Disordered Materials via Localized States

Electron transport via delocalized states above the mobility edge dominates the electrical conduction of disordered materials only at temperatures high enough to cause a significant fraction of the charge carriers fill these states. As the temperature decreases, the concentration of the electrons described by (9.9) decreases exponentially and so their contribution to electrical conductivity diminishes. Under these circumstances, tunneling transitions of electrons between localized states in the band tails dominate the charge transport in disordered semiconductors. This transport regime is called hopping conduction, since the incoherent sequence of tunneling transitions of charge carriers resembles a series of their hops between randomly distributed sites. Each site in this picture provides a spatially localized electron state with some energy ε . In the following we will assume that the localized states for electrons (concentration N_0) are randomly distributed in space and their energy distribution is described by the DOS function $g(\varepsilon)$:

$$g(\varepsilon) = \frac{N_0}{\varepsilon_0} G\left(\frac{\varepsilon}{\varepsilon_0}\right), \quad (9.22)$$

where ε_0 is the energy scale of the DOS distribution.

The tunneling transition probability of an electron from a localized state i to a localized state j that is lower in energy depends on the spatial separation r_{ij} between the sites i and j as

$$v_{ij}(r) = v_0 \exp\left(-\frac{2r_{ij}}{\alpha}\right), \quad (9.23)$$

where α is the localization length, which we assume to be equal for sites i and j . This length determines the exponential decay of the electron wavefunction in the localized states, as shown in Fig. 9.6. The pre-exponential factor v_0 in (9.23) depends on the electron interaction mechanism that causes the transition. Usually it is assumed that electron transitions contributing to charge transport in disordered materials are caused by interactions of electrons with phonons. Often the coefficient v_0 is simply assumed to be of the order of the phonon

In the following section we present the general concepts developed to describe electrical conduction in disordered solids at temperatures where tunneling transitions of electrons between localized states significantly contribute to charge transport.

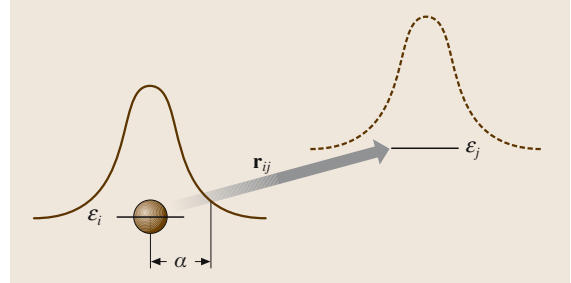


Fig. 9.6 Hopping transition between two localized states i and j with energies of ε_i and ε_j , respectively. The *solid* and *dashed* lines depict the carrier wavefunctions at sites i and j , respectively; α is the localization radius

frequency ($\approx 10^{13} \text{ s}^{-1}$), although a more rigorous approach is in fact necessary to determine v_0 . This should take into account the particular structure of the electron localized states and also the details of the interaction mechanism [9.29, 30].

When an electron transits from a localized state i to a localized state j that is higher in energy, the transition rate depends on the energy difference between the states. This difference is compensated for by absorbing a phonon with the corresponding energy [9.31]:

$$v_{ij}(r, \varepsilon_i, \varepsilon_j) = v_0 \exp\left(-\frac{2r_{ij}}{\alpha}\right) \times \exp\left(-\frac{\varepsilon_j - \varepsilon_i + |\varepsilon_j - \varepsilon_i|}{2k_B T}\right). \quad (9.24)$$

Equations (9.23) and (9.24) were written for the case in which the electron occupies site i whereas site j is empty. If the system is in thermal equilibrium, the occupation probabilities of sites with different energies are determined by Fermi statistics. This effect can be taken into account by modifying (9.24) and adding terms that account for the relative energy positions of sites i and

j with respect to the Fermi energy ε_F . Taking into account these occupation probabilities, one can write the transition rate between sites i and j in the form [9.31]

$$v_{ij} = v_0 \exp\left(-\frac{2r_{ij}}{a}\right) \times \exp\left(-\frac{|\varepsilon_i - \varepsilon_F| + |\varepsilon_j - \varepsilon_F| + |\varepsilon_j - \varepsilon_i|}{2k_B T}\right). \quad (9.25)$$

Using these formulae, the theoretical description of hopping conduction is easily formulated. One has to calculate the conductivity provided by transition events (the rates of which are described by (9.25)) in the manifold of localized states (where the DOS is described by (9.22)).

9.3.1 Nearest-Neighbor Hopping

Before presenting the correct solution to the hopping problem we would like to emphasize the following. The style of the theory for electron transport in disordered materials via localized states significantly differs from that used for theories of electron transport in ordered crystalline materials. While the description is usually based on various averaging procedures in crystalline systems, in disordered systems these averaging procedures can lead to extremely erroneous results. We believe that it is instructive to analyze some of these approaches in order to illustrate the differences between the descriptions of charge transport in ordered and disordered materials. To treat the scattering rates of electrons in ordered crystalline materials, one usually proceeds by averaging the scattering rates over the ensemble of scattering events. A similar procedure is often attempted for disordered systems too, although various textbooks (see, for instance, *Shklovskii and Efros* [9.14]) illustrate how erroneous such an approach can be in the case of disordered materials.

Let us consider the simplest example of hopping processes, namely the hopping of an electron through a system of isoenergetic sites randomly distributed in space with some concentration N_0 . It will be always assumed in this chapter that electron states are strongly localized and the strong inequality $N_0\alpha^3 \ll 1$ is fulfilled. In such a case the electrons prefer to hop between the spatially nearest sites and therefore this transport regime is often called nearest-neighbor hopping (NNH). This type of hopping transport takes place in many real systems at temperatures where the thermal energy $k_B T$ is larger than the energy scale of the DOS. In such sit-

uations the energy-dependent terms in (9.24) and (9.25) do not play any significant role and the hopping rates are determined solely by the spatial terms. The rate of transition of an electron between two sites i and j is described in this case by (9.23). The average transition rate is usually obtained by weighting this expression with the probability of finding the nearest neighbor at some particular distance r_{ij} , and by integrating over all possible distances:

$$\begin{aligned} \langle v \rangle &= \int_0^\infty dr v_0 \\ &\times \exp\left(-\frac{2r}{\alpha}\right) 4\pi r^2 N_0 \exp\left(-\frac{4\pi}{3} r^3 N_0\right) \\ &\approx \pi v_0 N_0 \alpha^3. \end{aligned} \quad (9.26)$$

Assuming that this average hopping rate describes the mobility, diffusivity and conductivity of charge carriers, one apparently comes to the conclusion that these quantities are linearly proportional to the density of localized states N_0 . However, experiments evidence an exponential dependence of the transport coefficients on N_0 .

Let us look therefore at the correct solution to the problem. This solution is provided in the case considered here, $N_0\alpha^3 \ll 1$, by percolation theory (see, for instance, *Shklovskii and Efros* [9.14]). In order to find the transport path, one connects each pair of sites if the relative separation between the sites is smaller than some given distance R , and checks whether there is a continuous path through the system via such sites. If such a path is absent, the magnitude of R is increased and the procedure is repeated. At some particular value $R = R_c$, a continuous path through the infinite system via sites with relative separations $R < R_c$ arises. Various mathematical considerations give the following relation for R_c [9.14]:

$$\frac{4\pi}{3} N_0 R_c^3 = B_c, \quad (9.27)$$

where $B_c = 2.7 \pm 0.1$ is the average number of neighboring sites available within a distance of less than R_c . The corresponding value of R_c should be inserted into (9.23) in order to determine kinetic coefficients such as the mobility, diffusivity and conductivity. The idea behind this procedure is as follows. Due to the exponential dependence of the transition rates on the distances between the sites, the rates for electron transitions over distances $r < R_c$ are much larger than those over distances R_c . Such fast transitions do not play any significant role as a limiting factor in electron transport and so they can

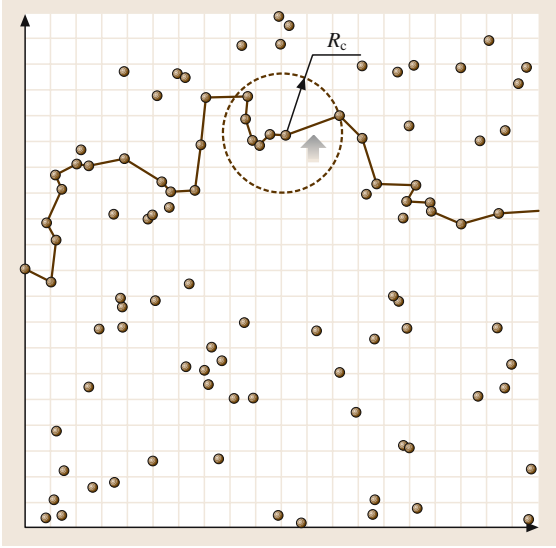


Fig. 9.7 A typical transport path with the lowest resistance. Circles depict localized states. The arrow points out the most “difficult” transition, with length R_c

be neglected in calculations of the resistivity of the system. Transitions over distances R_c are the slowest among those that are necessary for DC transport and hence such transitions determine the conductivity. The structure of the percolation cluster responsible for charge transport is shown schematically in Fig. 9.7. The transport path consists of quasi-one-dimensional segments, each containing a “difficult” transition over the distance $\approx R_c$. Using (9.23) and (9.27), one obtains the dependence of the conductivity on the concentration of localization sites in the form

$$\sigma = \sigma_0 \exp\left(-\frac{\gamma}{\alpha N_0^{1/3}}\right), \quad (9.28)$$

where σ_0 is the concentration-independent pre-exponential factor and $\gamma = 1.73 \pm 0.03$. Such arguments do not allow one to determine the exponent in the kinetic coefficients with an accuracy better than a number of the order of unity [9.14]. One should note that the quantity in the exponent in (9.28) is much larger than unity for a system with strongly localized states when the inequality $N_0 \alpha^3 \ll 1$ is valid. This inequality justifies the above derivation. The dependence described by (9.28) has been confirmed in numerous experimental studies of the hopping conductivity via randomly placed impurity atoms in doped crystalline semiconductors [9.14]. The drastic difference between this correct result and the erroneous one based on (9.26) is apparent. Unfortunately, the be-

lief of many researchers in the validity of the procedure based on the averaging of hopping rates is so strong that the agreement between (9.28) and experimental data is often called occasional. We would like to emphasize once more that the ensemble averaging of hopping rates leads to erroneous results. The magnitude of the average rate in (9.26) is dominated by rare configurations of very close pairs of sites with separations of the order of the localization length α . Of course, such pairs allow very fast electron transitions, but electrons cannot move over considerable distances using only these close pairs. Therefore the magnitude of the average transition rate is irrelevant for calculations of the hopping conductivity. The correct concentration dependence of the conductivity is given by (9.28). This result was obtained under the assumption that only spatial factors determine transition rates of electrons via localized states. This regime is valid at reasonably high temperatures.

If the temperature is not as high and the thermal energy $k_B T$ is smaller than the energy spread of the localized states involved in the charge transport process, the problem of calculating the hopping conductivity becomes much more complicated. In this case, the interplay between the energy-dependent and the distance-dependent terms in (9.24) and (9.25) determines the conductivity. The lower the temperature, the more important the energy-dependent terms in the expressions for transition probabilities of electrons in (9.24) and (9.25) become. If the spatially nearest-neighboring sites have very different energies, as shown in Fig. 9.8, the probability of an upward electron transition between these sites can be so low that it would be more favorable for this electron to hop to a more distant site at a closer energy. Hence the typical lengths of

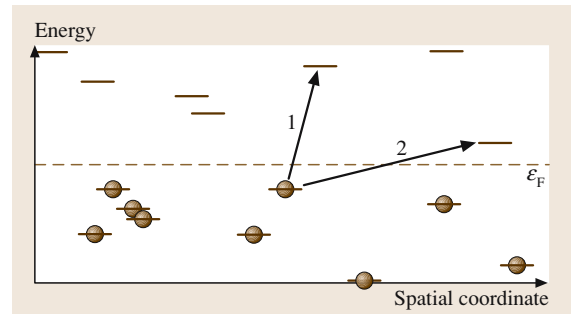


Fig. 9.8 Two alternative hopping transitions between occupied states (filled circles) and unoccupied states. The dashed line depicts the position of the Fermi level. Transitions (1) and (2) correspond to nearest-neighbor hopping and variable-range hopping regimes, respectively

electron transitions increase with decreasing temperature. This transport regime was termed “variable-range hopping”. Next we describe several useful concepts developed to describe this transport regime.

9.3.2 Variable-Range Hopping

The concept of variable-range hopping (VRH) was put forward by Mott (see *Mott and Davis* [9.32]) who considered electron transport via a system of randomly distributed localized states at low temperatures. We start by presenting Mott’s arguments. At low temperatures, electron transitions between states with energies in the vicinity of the Fermi level are most efficient for transport since filled and empty states with close energies can only be found in this energy range. Consider the hopping conductivity resulting from energy levels within a narrow energy strip with width $2\Delta\varepsilon$ symmetric to the Fermi level shown in Fig. 9.9. The energy width of the strip useful for electron transport can be determined from the relation

$$g(\varepsilon_F) \cdot \Delta\varepsilon \cdot r^3(\Delta\varepsilon) \approx 1. \quad (9.29)$$

This criterion is similar to that used in (9.27), although we do not care about numerical coefficients here. Here we have to consider the percolation problem in four-dimensional space since in addition to the spatial terms considered in Sect. 9.3.1 we now have to consider the energy too. The corresponding percolation problem for the transition rates described by (9.25) has not yet been solved precisely. In (9.29) it is assumed that the energy width $2\Delta\varepsilon$ is rather small and that the DOS function $g(\varepsilon)$ is almost constant in the range $\varepsilon_F \pm \Delta\varepsilon$. One can obtain

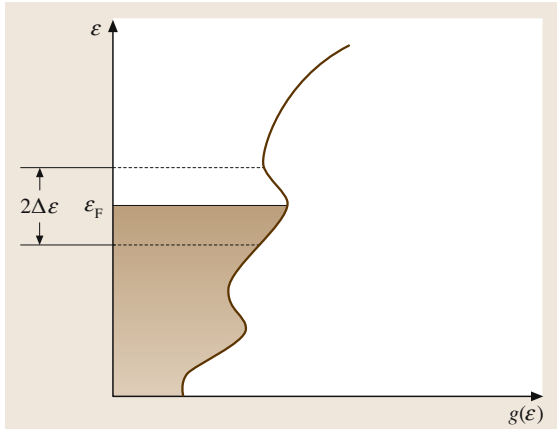


Fig. 9.9 Effective region in the vicinity of the Fermi level where charge transport takes place at low temperatures

the typical hopping distance from (9.29) as a function of the energy width $\Delta\varepsilon$ in the form

$$r(\Delta\varepsilon) \approx [g(\varepsilon_F)\Delta\varepsilon]^{-1/3}, \quad (9.30)$$

and substitute it into (9.24) in order to express the typical hopping rate

$$\nu = \nu_0 \exp\left(-\frac{2[g(\varepsilon_F)\Delta\varepsilon]^{-1/3}}{\alpha} - \frac{\Delta\varepsilon}{k_B T}\right). \quad (9.31)$$

The optimal energy width $\Delta\varepsilon$ that provides the maximum hopping rate can be determined from the condition $d\nu/d\Delta\varepsilon = 0$. The result reads

$$\Delta\varepsilon = \left(\frac{2k_B T}{3g^{1/3}(\varepsilon_F)}\right)^{3/4}. \quad (9.32)$$

After substitution of (9.32) into (9.31) one obtains Mott’s famous formula for temperature-dependent conductivity in the VRH regime

$$\sigma = \sigma_0 \exp\left[-\left(\frac{T_0}{T}\right)^{1/4}\right], \quad (9.33)$$

where T_0 is the characteristic temperature:

$$T_0 = \frac{\beta}{k_B g(\varepsilon_F) \alpha^3}. \quad (9.34)$$

Mott gave only a semi-quantitative derivation of (9.33), from which the exact value of the numerical constant β cannot be determined. Various theoretical studies in 3-D systems suggest values for β in the range $\beta = 10.0$ to $\beta = 37.8$. According to our computer simulations, the appropriate value is close to $\beta = 17.6$.

Mott’s law implies that the density of states in the vicinity of the Fermi level is energy-independent. However, it is known that long-range electron–electron interactions in a system of localized electrons cause a gap (the so-called Coulomb gap) in the DOS in the vicinity of the Fermi energy [9.33, 34]. The gap is shown schematically in Fig. 9.10. Using simple semiquantitative arguments, *Efros and Shklovskii* [9.33] suggested a parabolic shape for the DOS function

$$g(\varepsilon) = \frac{\eta \kappa^3}{e^6} (\varepsilon - \varepsilon_F)^2, \quad (9.35)$$

where κ is the dielectric constant, e is the elementary charge and η is a numerical coefficient. This result was later confirmed by numerous computer simulations (see, for example, *Baranovskii et al.* [9.35]). At low temperatures, the density of states near the Fermi level has a parabolic shape, and it vanishes exactly at the Fermi

energy. As the temperature rises, the gap disappears (see, for example, *Shlimak et al.* [9.36]).

As we have seen above, localized states in the vicinity of the Fermi energy are the most useful for transport at low temperatures. Therefore the Coulomb gap essentially modifies the temperature dependence of the hopping conductivity in the VRH regime at low temperatures. The formal analysis of the T -dependence of the conductivity in the presence of the Coulomb gap is similar to that for the Mott's law discussed above. Using the parabolic energy dependence of the DOS function, one arrives at the result

$$\sigma = \sigma_0 \exp \left[- \left(\frac{\tilde{T}_0}{T} \right)^{1/2} \right] \quad (9.36)$$

with $\tilde{T}_0 = \tilde{\beta} e^2 / (\kappa \alpha k_B)$, where $\tilde{\beta}$ is a numerical coefficient.

Equations (9.33) and (9.36) belong to the most famous theoretical results in the field of variable-range hopping conduction. However these formulae are usually of little help to researchers working with essentially noncrystalline materials, such as amorphous, vitreous or organic semiconductors. The reason is as follows. The above formulae were derived for the cases of either constant DOS (9.33) or a parabolic DOS (9.36) in the energy range associated with hopping conduction. These conditions can usually be met in the impurity band of a lightly doped crystalline semiconductor. In the most disordered materials, however, the energy distribution of the localized states is described by a DOS function that is very strongly energy-dependent. In amorphous, vitreous and microcrystalline semiconductors, the energy dependence of the DOS function is believed to be expo-

ponential, while in organic materials it is usually assumed to be Gaussian. In these cases, new concepts are needed in order to describe the hopping conduction. In the next section we present these new concepts and calculate the way the conductivity depends on temperature and on the concentration of localized states in various significantly noncrystalline materials.

9.3.3 Description of Charge-Carrier Energy Relaxation and Hopping Conduction in Inorganic Noncrystalline Materials

In most inorganic noncrystalline materials, such as vitreous, amorphous and polycrystalline semiconductors, the localized states for electrons are distributed over a rather broad energy range with a width of the order of an electronvolt. The DOS function that describes this energy distribution in such systems is believed to have a purely exponential shape

$$g(\varepsilon) = \frac{N_0}{\varepsilon_0} \exp \left(- \frac{\varepsilon}{\varepsilon_0} \right), \quad (9.37)$$

where the energy ε is counted positive from the mobility edge towards the center of the mobility gap, N_0 is the total concentration of localized states in the band tail, and ε_0 determines the energy scale of the tail. To be precise, we consider that electrons are the charge carriers here. The result for holes can be obtained in an analogous way. Values of ε_0 in inorganic noncrystalline materials are believed to vary between 0.025 eV and 0.05 eV, depending on the system under consideration.

It is worth noting that arguments in favor of a purely exponential shape for the DOS in the band tails of inorganic noncrystalline materials described by (9.37) cannot be considered to be well justified. They are usually based on a rather ambiguous interpretation of experimental data. One of the strongest arguments in favor of (9.37) is the experimental observation of the exponential decay of the light absorption coefficient for photons with an energy deficit ε with respect to the energy width of the mobility gap (see, for example, *Mott and Davis* [9.32]). One should mention that this argument is valid only under the assumption that the energy dependence of the absorption coefficient is determined solely by the energy dependence of the DOS. However, in many cases the matrix element for electron excitation by a photon in noncrystalline materials also strongly depends on energy [9.14, 37]. Hence any argument for the shape of the DOS based on the energy dependence of the light absorption coefficient should be taken very cautiously. Another argument in favor of (9.37) comes

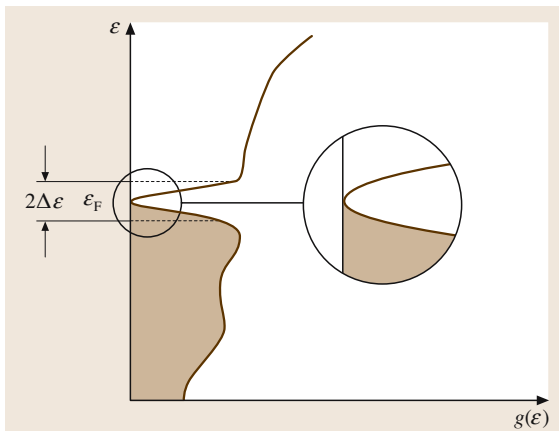


Fig. 9.10 Schematic view of the Coulomb gap. The insert shows the parabolic shape of the DOS near the Fermi level

from the measurements of dispersive transport in time-of-flight experiments. In order to interpret the observed time dependence of the mobility of charge carriers, one usually assumes that the DOS for the band tail takes the form of (9.37) (see, for example, *Orenstein and Kastner* [9.38]). One of the main reasons for such an assumption is probably the ability to solve the problem analytically without elaborate computer work.

In the following we start our consideration of the problem by also assuming that the DOS in a band tail of a noncrystalline material has an energy dependence that is described by (9.37). This simple function will allow us to introduce some valuable concepts that have been developed to describe dynamic effects in noncrystalline materials in the most transparent analytical form. We first present the concept of the so-called transport energy, which, in our view, provides the most transparent description of the charge transport and energy relaxation of electrons in noncrystalline materials.

The Concept of the Transport Energy

The crucial role of a particular energy level in the hopping transport of electrons via localized band-tail states with the DOS described by (9.37) was first recognized by *Grünewald and Thomas* [9.39] in their numerical analysis of equilibrium variable-range hopping conductivity. This problem was later considered by *Shapiro and Adler* [9.40], who came to the same conclusion as *Grünewald and Thomas*, namely that the vicinity of one particular energy level dominates the hopping transport of electrons in the band tails. In addition, they achieved an analytical formula for this level and showed that its position does not depend on the Fermi energy.

Independently, the rather different problem of nonequilibrium energy relaxation of electrons by hopping through the band tail with the DOS described by (9.37) was solved at the same time by *Monroe* [9.41]. He showed that, starting from the mobility edge, an electron most likely makes a series of hops downward in energy. The manner of the relaxation process changes at some particular energy ε_t , which Monroe called the transport energy (TE). The hopping process near and below TE resembles a multiple-trapping type of relaxation, with the TE playing a role similar to the mobility edge. In the multiple-trapping relaxation process [9.38], only electron transitions between delocalized states above the mobility edge and the localized band-tail states are allowed, while hopping transitions between the localized tail states are neglected. Hence, every second transition brings the electron to the mobility edge. The TE of *Monroe* [9.41] coincides exactly with the energy

level discovered by *Grünewald and Thomas* [9.39] and by *Shapiro and Adler* [9.40] for equilibrium hopping transport.

Shklovskii et al. [9.42] have shown that the same energy level ε_t also determines the recombination and transport of electrons in the nonequilibrium steady state under continuous photogeneration in a system with the DOS described by (9.37).

It is clear, then, that the TE determines both equilibrium and nonequilibrium and both transient and steady-state transport phenomena. The question then arises as to why this energy level is so universal that electron hopping in its vicinity dominates all transport phenomena. Below we derive the TE by considering a single hopping event for an electron localized deep in the band tail. It is the transport energy that maximizes the hopping rate as a final electron energy in the hop, independent of its initial energy [9.43]. All derivations below are carried out for the case $k_B T < \varepsilon_0$.

Consider an electron in a tail state with energy ε_i . According to (9.24), the typical rate of downward hopping of such an electron to a neighboring localized state deeper in the tail with energy $\varepsilon_j \geq \varepsilon_i$ is

$$v_{\downarrow}(\varepsilon_i) = v_0 \exp\left(-\frac{2r(\varepsilon_i)}{\alpha}\right), \quad (9.38)$$

where

$$r(\varepsilon) \approx \left[\frac{4\pi}{3} \int_{\varepsilon_i}^{\infty} g(x) dx \right]^{-1/3}. \quad (9.39)$$

The typical rate of upward hopping for such an electron to a state less deep in the tail with energy $\varepsilon_j \leq \varepsilon_i$ is

$$v_{\uparrow}(\varepsilon_i, \delta) = v_0 \exp\left[-\frac{2r(\varepsilon_i - \delta)}{\alpha} - \frac{\delta}{k_B T}\right], \quad (9.40)$$

where $\delta = \varepsilon_i - \varepsilon_j \geq 0$. This expression is not exact. The average nearest-neighbor distance, $r(\varepsilon_i - \delta)$, is based on all states deeper than $\varepsilon_i - \delta$. For the exponential tail, this is equivalent to considering a slice of energy with a width of the order ε_0 . This works for a DOS that varies slowly compared with $k_B T$, but not in general. It is also assumed for simplicity that the localization length, α , does not depend on energy. The latter assumption can be easily jettisoned at the cost of somewhat more complicated forms of the following equations.

We will analyze these hopping rates at a given temperature T , and try to find the energy difference δ that provides the fastest typical hopping rate for an electron placed initially at energy ε_i . The corresponding energy

difference, δ , is determined by the condition

$$\frac{d\nu_{\uparrow}(\varepsilon_i, \delta)}{d\delta} = 0. \quad (9.41)$$

Using (9.37), (9.39) and (9.40), we find that the hopping rate in (9.40) has its maximum at

$$\delta = \varepsilon_i - 3\varepsilon_0 \ln \frac{3\varepsilon_0(4\pi/3)^{1/3} N_0^{1/3} \alpha}{2k_B T}. \quad (9.42)$$

The second term in the right-hand side of (9.42) is called the transport energy ε_t after *Monroe* [9.41]:

$$\varepsilon_t = 3\varepsilon_0 \ln \frac{3\varepsilon_0(4\pi/3)^{1/3} N_0^{1/3} \alpha}{2k_B T}. \quad (9.43)$$

We see from (9.42) that the fastest hop occurs to the state in the vicinity of the TE, independent of the initial energy ε_i , provided that ε_i is deeper in the tail than ε_t ; in other words, if $\delta \geq 0$. This result coincides with that of *Monroe* [9.41]. At low temperatures, the TE ε_t is situated deep in the band tail, and as the temperature rises it moves upward towards the mobility edge. At some temperature T_c , the TE merges with the mobility edge. At higher temperatures, $T > T_c$, the hopping exchange of electrons between localized band tail states becomes inefficient and the dynamic behavior of electrons is described by the well-known multiple-trapping model (see, for instance, *Orenstein and Kastner* [9.38]). At low temperatures, $T < T_c$, the TE replaces the mobility edge in the multiple-trapping process [9.41], as shown in Fig. 9.11. The width, W , of the maximum of the hopping rate is determined by the requirement that near ε_t the hopping rate, $\nu_{\uparrow}(\varepsilon_i, \delta)$, differs by less than a factor of e from the value $\nu_{\uparrow}(\varepsilon_t, \varepsilon_t - \varepsilon_t)$. One finds [9.42]

$$W = \sqrt{6\varepsilon_0 k_B T}. \quad (9.44)$$

For shallow states with $\varepsilon_i \leq \varepsilon_t$, the fastest hop (on average) is a downward hop to the nearest spatially localized state in the band tail, with the rate determined by (9.38) and (9.39). We recall that the energies of electron states are counted positive downward from the mobility edge towards the center of the mobility gap. This means that electrons in the shallow states with $\varepsilon_i \leq \varepsilon_t$ normally hop into deeper states with $\varepsilon > \varepsilon_t$, whereas electrons in the deep states with $\varepsilon_i > \varepsilon_t$ usually hop upward in energy into states near ε_t in the energy interval W , determined by (9.44).

This shows that ε_t must play a crucial role in those phenomena, which are determined by electron hopping in the band tails. This is indeed the case, as shown in numerous review articles where comprehensive theories based on the concept of the TE can be found (see,

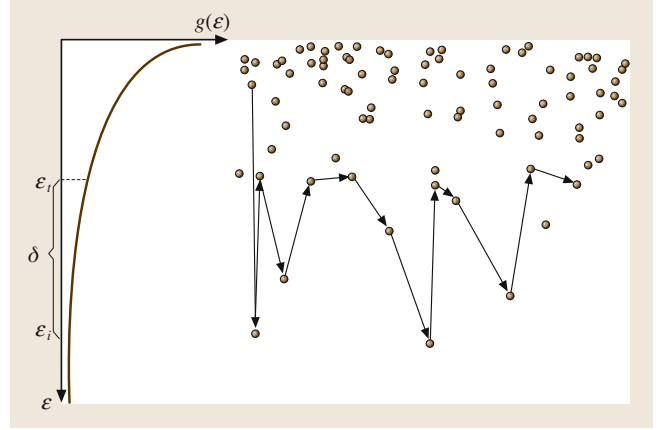


Fig. 9.11 Hopping path via the transport energy. In the left frame, the exponential DOS is shown schematically. The right frame depicts the transport path constructed from upward and downward hops. The upward transitions bring the charge carrier to sites with energies close to the transport energy ε_t

for instance, *Shklovskii et al.* [9.42]). We will consider only one phenomenon here for illustration, namely the hopping energy relaxation of electrons in a system with the DOS described by (9.37). This problem was studied initially by *Monroe* [9.41].

Consider an electron in some localized shallow energy state close to the mobility edge. Let the temperature be low, $T < T_c$, so that the TE, ε_t , lies well below the mobility edge, which has been chosen here as a reference energy, $\varepsilon = 0$. The aim is to find the typical energy, $\varepsilon_d(t)$, of our electron as a function of time, t . At early times, as long as $\varepsilon_d(t) < \varepsilon_t$, the relaxation is governed by (9.38) and (9.39). The depth $\varepsilon_d(t)$ of an electron in the band tail is determined by the condition

$$\nu_{\downarrow}[\varepsilon_d(t)] t \approx 1. \quad (9.45)$$

This leads to the double logarithmical dependence $\varepsilon_d(t) \propto \varepsilon_0 \ln[\ln(\nu_0 t)] + C$, where constant C depends on $\varepsilon_0, N_0, \alpha$ in line with (9.38) and (9.39). Indeed, (9.38) and (9.45) prescribe the logarithmic form of the time dependence of the hopping distance, $r(t)$, and (9.37) and (9.39) then lead to another logarithmic dependence $\varepsilon_d[r(t)]$ [9.41]. At the time

$$t_c \approx \nu_0^{-1} \exp\left(\frac{3\varepsilon_0}{k_B T}\right) \quad (9.46)$$

the typical electron energy, $\varepsilon_d(t)$, approaches the TE ε_t , and the style of the relaxation process changes. At $t > t_c$, every second hop brings the electron into states in the vicinity of the TE ε_t from where it falls downward in

energy to the nearest (in space) localization site. In the latter relaxation process, the typical electron energy is determined by the condition [9.41]

$$\nu_{\uparrow}[\varepsilon_d(t), \varepsilon_t] t \approx 1, \quad (9.47)$$

where $\nu_{\uparrow}[\varepsilon_d(t), \varepsilon_t]$ is the typical rate of electron hopping upward in energy toward the TE [9.41]. This condition leads to a typical energy position of the relaxing electron at time t of

$$\varepsilon_d(t) \approx 3\varepsilon_0 \ln[\ln(\nu_0 t)] - \varepsilon_0 \left[8 / (N_0 \alpha^3) \right]. \quad (9.48)$$

This is a very important result, which shows that in a system where the DOS has a pure exponential energy dependence, described by (9.37), the typical energy of a set of independently relaxing electrons would drop deeper and deeper into the mobility gap with time. This result is valid as long as the electrons do not interact with each other, meaning that the occupation probabilities of the electron energy levels are not taken into account. This condition is usually met in experimental studies of transient processes, in which electrons are excited by short (in time) pulses, which are typical of time-of-flight studies of the electron mobility in various disordered materials. In this case, only a small number of electrons are present in the band tail states. Taking into account the huge number of localized band tail states in most disordered materials, one can assume that most of the states are empty and so the above formulae for the hopping rates and electron energies can be used. In this case the electron mobility is a time-dependent quantity [9.41]. A transport regime in which mobility of charge carriers is time-dependent is usually called dispersive transport (see, for example, *Mott and Davis* [9.32], *Orenstein and Kastner* [9.38], *Monroe* [9.41]). Hence we have to conclude that the transient electron mobility in inorganic noncrystalline materials with the DOS in the band tails as described by (9.37) is a time-dependent quantity and the transient electrical conductivity has dispersive character. This is due to the nonequilibrium behavior of the charge carriers. They continuously drop in energy during the course of the relaxation process.

In some theoretical studies based on the Fokker–Planck equation it has been claimed that the maximum of the energy distribution of electrons coincides with the TE ε_t and hence it is independent of time. This statement contradicts the above result where the maximum of the distribution is at energy $\varepsilon_d(t)$, given by (9.48). The Fokker–Planck approach presumes the diffusion of

charge carriers over energy. Hence it is invalid for describing the energy relaxation in the exponential tails, in which electron can move over the full energy width of the DOS (from a very deep energy state toward the TE) in a single hopping event.

In the equilibrium conditions, when electrons in the band tail states are provided by thermal excitation from the Fermi energy, a description of the electrical conductivity can easily be derived using (9.5)–(9.7) [9.39]. The maximal contribution to the integral in (9.5) comes from the electrons with energies in the vicinity of the TE ε_t , in an energy range with a width, W , described by (9.44). Neglecting the temperature dependence of the pre-exponential factor, σ_0 , one arrives at the temperature dependence of the conductivity:

$$\sigma \approx \sigma_0 \exp \left(-\frac{2r(\varepsilon_t)}{B_c^{-1/3} \alpha} - \frac{\varepsilon_F - \varepsilon_t}{k_B T} \right), \quad (9.49)$$

where coefficient $B_c \approx 2.7$ is inserted in order to take into account the need for a charge carrier to move over macroscopic percolation distances in order to provide low-frequency charge transport.

A very similar theory is valid for charge transport in noncrystalline materials under stationary excitation of electrons (for example by light) [9.42]. In such a case, one first needs to develop a theory for the steady state of the system under stationary excitation. This theory takes into account various recombination processes for charge carriers and provides their stationary concentration along with the position of the quasi-Fermi energy. After solving this recombination problem, one can follow the track of the theory of charge transport in quasi-thermal equilibrium [9.39] and obtain the conductivity in a form similar to (9.49), where ε_F is the position of the quasi-Fermi level. We skip the corresponding (rather sophisticated) formulae here. Interested readers can find a comprehensive description of this sort of theory for electrical conductivity in the literature (see, for instance, *Shklovskii et al.* [9.42]).

Instead, in the next section we will consider a very interesting problem related to the nonequilibrium energy relaxation of charge carriers in the band tail states. It is well known that at low temperatures, $T \leq 50$ K, the photoconductivities of various inorganic noncrystalline materials, such as amorphous and microcrystalline semiconductors, do not depend on temperature [9.44–46]. At low temperatures, the TE ε_t lies very deep in the band tail and most electrons hop downward in energy, as described by (9.38) and (9.39). In such a regime, the

photoconductivity is a temperature-independent quantity determined by the loss of energy during the hopping of electrons via the band-tail states [9.47]. During this hopping relaxation, neither the diffusion coefficient D nor the mobility of the carriers μ depend on temperature, and the conventional form of Einstein's relationship $\mu = eD/k_B T$ cannot be valid. The question then arises as to what the relation between μ and D is for hopping relaxation. We answer this question in the following section.

Einstein's Relationship for Hopping Electrons

Let us start by considering a system of nonequilibrium electrons in the band tail states at $T = 0$. The only process that can happen with an electron is its hop downward in energy (upward hops are not possible at $T = 0$) to the nearest localized state in the tail. Such a process is described by (9.37)–(9.39). If the spatial distribution of localized tail states is isotropic, the probability of finding the nearest neighbor is also isotropic in the absence of the external electric field. In this case, the process of the hopping relaxation of electrons resembles diffusion in space. However, the median length of a hop (the distance r to the nearest available neighbor), as well as the median time, $\tau = v_{\downarrow}^{-1}(r)$, of a hop [see (9.38)] increases during the course of relaxation, since the hopping process brings electrons deeper into the tail. Nevertheless, one can ascribe a diffusion coefficient to such a process [9.42]:

$$D(r) = \frac{1}{6} v_{\downarrow}(r) r^2. \quad (9.50)$$

Here $v_{\downarrow}(r)r^2$ replaces the product of the “mean free path” r and the “velocity” $r \cdot v_{\downarrow}(r)$, and the coefficient $1/6$ accounts for the spatial symmetry of the problem. According to (9.37)–(9.39) and (9.50), this diffusion coefficient decreases exponentially with increasing r and hence with the number of successive electron hops in the relaxation process.

In order to calculate the mobility of electrons during hopping relaxation under the influence of the electric field, one should take into account the spatial asymmetry of the hopping process due to the field [9.47, 48]. Let us consider an electron in a localized state at energy ε . If an external electric field with a strength F is applied along direction x , the concentration of tail states available to this hopping electron at $T = 0$ (in other words those that have energies deeper in the tail than ε) is [9.47]

$$N(\varepsilon, x) = N(\varepsilon) \left(1 + \frac{eFx}{\varepsilon_0} \right), \quad (9.51)$$

where

$$N(\varepsilon) = \int_{\varepsilon}^{\infty} g(\varepsilon) d\varepsilon = N_0 \exp \left(-\frac{\varepsilon}{\varepsilon_0} \right). \quad (9.52)$$

It was assumed in the derivation of (9.51) that $eFx \ll \varepsilon_0$.

Due to the exponential dependence of the hopping rate on the hopping length r , the electron predominantly hops to the nearest tail state among the available states if $r \gg \alpha$, which we assume to be valid. Let us calculate the average projection $\langle x \rangle$ on the field direction of the vector \mathbf{r} from the initial states at energy ε to the nearest available neighbor among sites with a concentration $N(\varepsilon, x)$ determined by (9.51). Introducing spherical coordinates with the angle θ between r and the x -axis, we obtain [9.48]

$$\begin{aligned} \langle x \rangle &= \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin \theta \\ &\times \int_0^{\infty} [dr \cdot r^3 \cos(\theta) \cdot N(\varepsilon, r \cos \theta)] \\ &\times \exp \left[-\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin \theta \right. \\ &\times \left. \int_0^r dr' r'^2 N(\varepsilon, r' \cos \theta) \right]. \end{aligned} \quad (9.53)$$

Substituting (9.51) for $N(\varepsilon, r \cos \theta)$, calculating the integrals in (9.53) and omitting the second-order terms

$$\left(\frac{eN^{-1/3}(\varepsilon)F}{\varepsilon_0} \right)^2 \ll 1, \quad (9.54)$$

we obtain

$$\langle x \rangle = \frac{eFN^{-2/3}(\varepsilon)}{3\varepsilon_0} \frac{\Gamma(5/3)}{(4\pi/3)^{2/3}}, \quad (9.55)$$

where Γ is the gamma-function and $N(\varepsilon)$ is determined by (9.52). Equation (9.55) gives the average displacement in the field direction of an electron that hops downward from a state at energy ε to the nearest available neighbor in the band tail. The average length $\langle r \rangle$ of

such a hop is

$$\begin{aligned} \langle r \rangle &= \int_0^\infty dr 4\pi r^3 N(\varepsilon) \exp \left[-\frac{4\pi}{3} N(\varepsilon) r^3 \right] \\ &= \left(\frac{4\pi N(\varepsilon)}{3} \right)^{-1/3} \Gamma \left(\frac{4}{3} \right). \end{aligned} \quad (9.56)$$

One can ascribe to the hopping process a mobility

$$\begin{aligned} \mu &= \frac{v}{F} = \frac{\langle x \rangle v(\langle r \rangle)}{F} \\ &= \frac{e N^{-2/3}(\varepsilon) v(\langle r \rangle)}{3\varepsilon_0} \frac{\Gamma(5/3)}{(4\pi/3)^{2/3}} \end{aligned} \quad (9.57)$$

and a diffusion coefficient

$$\begin{aligned} D &= \frac{1}{6} \langle r \rangle^2 v(\langle r \rangle) \\ &= \frac{1}{6} N^{-2/3}(\varepsilon) v(\langle r \rangle) \frac{\Gamma^2(4/3)}{(4\pi/3)^{2/3}}. \end{aligned} \quad (9.58)$$

Expressions (9.57) and (9.58) lead to a relationship between μ and D of the form

$$\mu = \frac{2\Gamma(5/3)}{\Gamma^2(4/3)} \frac{e}{\varepsilon_0} D \approx 2.3 \frac{e}{\varepsilon_0} D. \quad (9.59)$$

This formula replaces the Einstein's relationship $\mu = eD/k_B T$ for electron hopping relaxation in the exponential band tail. Several points should be noted about this result. First of all, one should clearly realize that (9.59) is valid for nonequilibrium energy-loss relaxation in which only downward (in energy) transitions between localized states can occur. This regime is valid only at low temperatures when the TE ε_t is very deep in the band tail. As the temperature increases, the upward hops become more and more efficient for electron relaxation. Under these circumstances, the relation between μ and D evolves gradually with rising temperature from its temperature-independent form at $T = 0$ to the conventional Einstein's relationship, $\mu = eD/k_B T$ [9.50, 51]. Secondly, one should realize that (9.59) was derived in the linear regime with respect to the applied field under the assumption that $eFx \ll \varepsilon_0$. According to (9.55), the quantity $\langle x \rangle$ is proportional to $N^{-2/3}(\varepsilon) = N_0^{-2/3} \exp [2\varepsilon/(3\varepsilon_0)]$, in other words it increases exponentially during the course of the relaxation toward larger localization energies ε . This means that for deep localized states in the band tail, the condition $eFx \ll \varepsilon_0$ breaks down. The boundary energy for application of the linear theory depends on the strength of the electric field, F . As F decreases, this boundary energy

drops deeper into the tail. However, for any F , there is always a boundary energy in the tail below which the condition $eFx \ll \varepsilon_0$ cannot be fulfilled and where nonlinear effects play the decisive role in the hopping conduction of charge carriers. In the next section we show how one can describe these nonlinear effects with respect to the applied electric field.

Nonlinear Effects in Hopping Conduction

Transport phenomena in inorganic noncrystalline materials, such as amorphous semiconductors, under the influence of high electric fields are the foci for intensive experimental and theoretical study. This is due to observations of strong nonlinearities in the dependencies of the dark conductivity [9.11, 52, 53], the photoconductivity [9.49] and the charge carrier drift mobility [9.54–56] on the field for high electric fields. These effects are most pronounced at low temperatures, when charge transport is determined by electron hopping via localized band tail states (Fig. 9.12).

Whereas the field-dependent hopping conductivity at low temperatures has always been a challenge to describe theoretically, theories for the temperature dependence of the hopping conductivity in low electric fields have been successfully developed for all of the transport regimes discussed: for the dark conductivity [9.39], for the drift mobility [9.41], and for the photoconductivity [9.42]. In all of these theories, hopping transitions of electrons between localized states in the exponential band tails play a decisive role, as described above in (9.37)–(9.59).

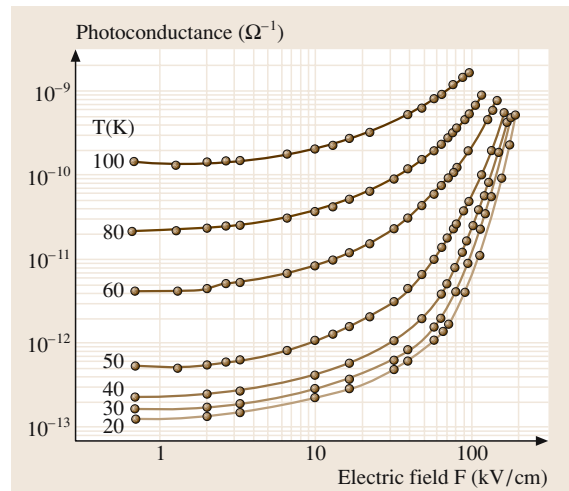


Fig. 9.12 Dependence of the photoconductivity in a-Si:H on the electric field at different temperatures [9.49]

Shklovskii [9.57] was the first to recognize that a strong electric field plays a similar role to that of temperature in hopping conduction. In order to obtain the field dependence of the conductivity $\sigma(F)$ at high fields, *Shklovskii* [9.57] replaced the temperature T in the well-known dependence $\sigma(T)$ for low fields by a function $T_{\text{eff}}(F)$ of the form

$$T_{\text{eff}} = \frac{eF\alpha}{2k_B}, \quad (9.60)$$

where e is the elementary charge, k_B is the Boltzmann constant, and α is the localization length of electrons in the band tail states. A very similar result was obtained later by *Grünwald* and *Movaghar* [9.58] in their study of the hopping energy relaxation of electrons through band tails at very low temperatures and high electric fields. The same idea was also used by *Shklovskii* et al. [9.42], who suggested that, at $T = 0$, one can calculate the field dependence of the stationary photoconductivity in amorphous semiconductors by replacing the laboratory temperature T in the formulae of the low-field finite-temperature theory by an effective temperature $T_{\text{eff}}(F)$ given by (9.60).

It is easy to understand why the electric field plays a role similar to that of temperature in the energy relaxation of electrons. Indeed, in the presence of the field, the number of sites available at $T = 0$ is significantly enhanced in the field direction, as shown in Fig. 9.13. Hence electrons can relax faster at higher fields. From the figure it is apparent that an electron can increase its energy with respect to the mobility edge by an amount $\varepsilon = eFx$ in a hopping event over a distance x in the direction prescribed by the electric field. The process is reminiscent of thermal activation. The analogy becomes tighter when we express the transition rate for this hop as

$$\begin{aligned} \nu &= \nu_0 \exp\left(-\frac{2x}{\alpha}\right) = \nu_0 \exp\left(-\frac{2\varepsilon}{eF\alpha}\right) \\ &= \nu_0 \exp\left(-\frac{\varepsilon}{k_B T_{\text{eff}}(F)}\right), \end{aligned} \quad (9.61)$$

where $T_{\text{eff}}(F)$ is provided by (9.60).

This electric field-induced activation at $T = 0$ produces a Boltzmann tail to the energy distribution function of electrons in localized states as shown by numerical calculations [9.59, 60]. In Fig. 9.12, the field-dependent photoconductivity in a-Si:H is shown for several temperatures [9.49]. If we compare the photoconductivity at the lowest measured temperature,

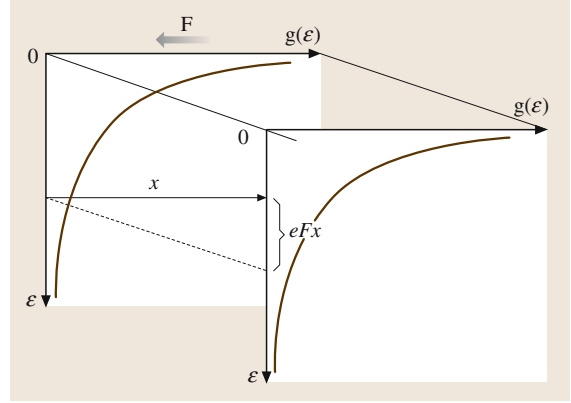


Fig. 9.13 Tunneling transition of a charge carrier in the band tail that is affected by a strong electric field. Upon traveling the distance x , the carrier acquires the energy eFx , where F is the strength of the electric field, and e is the elementary charge

$T = 20$ K in Fig. 9.12, with the low-field photoconductivity at $T = T_{\text{eff}} = \frac{eF\alpha}{2k_B}$ as measured by *Hoheisel* et al. [9.44] and by *Stradins* and *Fritzsche* [9.45], we come to the conclusion that the data agree quantitatively if one assumes that the localization length $\alpha = 1.05$ nm [9.42], which is very close to the value $\alpha \approx 1.0$ nm found for a-Si:H from independent estimates [9.11]. This comparison shows that the concept of the effective temperature based on (9.60) provides a powerful tool for estimating transport coefficient non-linearity with respect to the electric field using the low-field results for the temperature dependencies of such coefficients.

However, experiments are usually carried out not at $T = 0$ but at finite temperatures, and so the question of how to describe transport phenomena in the presence of both factors, finite T and high F , arises. By studying the steady state energy distribution of electrons in numerical calculations and computer simulations [9.59, 60], as well as straightforward computer simulations of the steady-state hopping conductivity and the transient energy relaxation of electrons [9.61], the following result was found. The whole set of transport coefficients can be represented by a function with a single parameter $T_{\text{eff}}(F, T)$

$$T_{\text{eff}}(F, T) = \left[T^\beta + \left(\gamma \frac{eF\alpha}{k_B} \right)^\beta \right]^{1/\beta}, \quad (9.62)$$

where $\beta \approx 2$ and γ is between 0.5 and 0.9 depending on which transport coefficient is considered [9.61]. We

are aware of no analytical theory that can support this numerical result.

To wrap up this section we would like to make the following remark. It is commonly claimed in the scientific literature that transport coefficients in the hopping regime should have a purely exponential dependence on the applied electric field. The idea behind such statements seems rather transparent. Electric field diminishes potential barriers between localized states by an amount $\Delta\varepsilon = eFx$, where x is the projection of the hopping radius on the field direction. The field should therefore diminish the activation energies in (9.24) and (9.25) by this amount, leading to the term $\exp(eFx/k_B T)$ in the expressions for the charge carrier mobility, diffusivity and conductivity. One should, however, take into account that hopping transport in all real materials is essentially described by the variable-range hopping process. In such a process, as discussed above, the interplay between spatial and energy-dependent terms in the exponents of the transition probabilities determine the conduction path. Therefore it is not enough to solely take into account the influence of the strong electric field on the activation energies of single hopping transitions. One should consider the modification of the whole transport path due to the effect of the strong field. It is this **VRH** nature of the hopping process that leads to a more complicated field dependence for the transport coefficients expressed by (9.60)–(9.62).

We have now completed our description of electron transport in inorganic disordered materials with exponential **DOS** in the band tails. In the next section we tackle the problem of charge transport in organic disordered materials.

9.3.4 Description of Charge Carrier Energy Relaxation and Hopping Conduction in Organic Noncrystalline Materials

Electron transport and energy relaxation in disordered organic solids, such as molecularly doped polymers, conjugated polymers and organic glasses, has been the subject of intensive experimental and theoretical study for more than 20 years. Although there is a wide array of different disordered organic solids, the charge transport process is similar in most of these materials. Even at the beginning of the 1980s it was well understood that the main transport mechanism in disordered organic media is the hopping of charge carriers via spatially randomly distributed localized states. Binary systems like doped polymeric matrices provide canonical examples of disordered organic materials that exhibit

the hopping transport mechanism. Examples include polyvinylcarbazole (**PVK**) or bis-polycarbonate (**Lexan**) doped with either strong electron acceptors such as trinitrofluorenone acting as an electron transporting agent, or strong electron donors such as derivatives of triphenylamine of triphenylmethane for hole transport [9.62, 63]. To avoid the need to specify whether transport is carried by electrons or holes each time, we will use a general notation of “charge carrier” below. The results are valid for both types of carrier – electrons or holes. Charge carriers in disordered organic materials are believed to be strongly localized [9.18, 62–64]. The localization centers are molecules or molecular subunits, henceforth called sites. These sites are located in statistically different environments. As a consequence, the site energies, which are to great extent determined by electronic polarization, fluctuate from site to site. The fluctuations are typically on the order of 0.1 eV [9.65]. This is about one order of magnitude larger than the corresponding transfer integrals [9.65]. Therefore carrier wavefunctions can be considered to be strongly localized [9.65].

As discussed above, the crucial problem when developing a theoretical picture for hopping transport is the structure of the energy spectrum of localized states, **DOS**. It is believed that, unlike inorganic noncrystalline materials where the **DOS** is believed exponential, the energy dependence of the **DOS** in organic disordered solids is Gaussian (see *Bässler* [9.18] and references therein),

$$g(\varepsilon) = \frac{N_0}{\varepsilon_0 \sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2\varepsilon_0^2}\right), \quad (9.63)$$

where N_0 is the total concentration of states and ε_0 is the energy scale of the **DOS**. The strongest evidence in favor of such an energy spectrum in disordered organic materials is the ability to reproduce the observed experimentally temperature dependence of the carrier mobility and that of hopping conductivity assuming the Gaussian **DOS** in computer simulations [9.18, 66]. It has been observed in numerous experimental studies [9.67–73] that the temperature dependence of the drift mobility of charge carriers in disordered organic solids takes the form

$$\mu \propto \exp\left[-\left(\frac{T_0}{T}\right)^2\right] \quad (9.64)$$

with a characteristic temperature T_0 , as shown in Fig. 9.14a. Computer simulations and theoretical calculations [9.65, 66, 74, 75] with the Gaussian **DOS**

described by (9.63) give a dependence of the form

$$\mu \propto \exp \left[- \left(C \frac{\varepsilon_0}{k_B T} \right)^2 \right], \quad (9.65)$$

where C is a numerical coefficient. Computer simulations [9.65, 66] give a value $C \approx 0.69$ for this coefficient, and analytical calculations [9.74, 75] predict a similar value of $C \approx 0.64$. Equation (9.65) is often used to determine the parameter ε_0 of the DOS from experimental measurements of the $\ln(\mu)$ versus $(1/T)^2$ dependences (see, for example, Ochse et al. [9.71]).

One may wonder whether the theoretical description of hopping conduction and carrier energy relaxation in a system with a Gaussian DOS (9.63) should differ significantly from the theory described above for disordered

systems with a purely exponential DOS (9.37). The answer to this question is yes. The reason becomes clear if one considers the behavior of a single charge carrier in an empty system. In an empty system with an exponential DOS, a charge carrier always (on average) falls downward in energy if $k_B T < \varepsilon_0$ [see (9.45)–(9.48)], and its mobility continuously decreases with time; however, in a system with a Gaussian DOS, a particular energy level ε_∞ determines the equilibrium energy position of a charge carrier. When it is located at some site with high energy in the Gaussian DOS, the charge carrier first hops via localized states so that its average energy $\varepsilon_d(t)$ decreases until it achieves the energy level ε_∞ after some typical time period τ_{rel} . At times $t < \tau_{\text{rel}}$ the behavior of the carrier qualitatively resembles that seen for the purely exponential DOS. The downward hops are then replaced by relaxation hops that send the carrier upward to the transport energy, and the carrier mobility at $t < \tau_{\text{rel}}$ decreases with time. However, in contrast with the case for the exponential DOS, in a Gaussian DOS the carrier mobility becomes time-independent after a time τ_{rel} , when the average carrier energy reaches the level ε_∞ . At $t > \tau_{\text{rel}}$, the dispersive transport regime with time-dependent carrier mobility is replaced by a quasi-equilibrium so-called “Gaussian transport” regime, in which the spatial spreading of the carrier packet with time can be described by the traditional diffusion picture with a time-independent diffusion coefficient.

The peculiarity of the hopping energy relaxation of charge carriers in a system with a Gaussian DOS described above makes it easier to describe charge transport at times $t > \tau_{\text{rel}}$ than in the case of the exponential DOS. In the latter case, only the presence of a significant number of carriers in a quasi Fermi level can make kinetic coefficients such as mobility, diffusivity and conductivity time-independent and hence conventionally measurable and discussible quantities. In the case of the Gaussian DOS, these kinetic coefficients are not time-dependent at times $t > \tau_{\text{rel}}$. Moreover, in diluted systems one can calculate these coefficients by considering the behavior of a single charge carrier. This makes theoretical considerations of electrical conductivity in organic disordered solids with a Gaussian DOS much easier than when considering inorganic noncrystalline materials with an exponential DOS. Let us now calculate ε_∞ , τ_{rel} and μ in disordered organic solids with a Gaussian DOS.

Computer simulations [9.66] and analytical calculations [9.74, 75] show that the mean energy of the independently hopping carriers, initially distributed randomly over all states in the Gaussian DOS, decreases

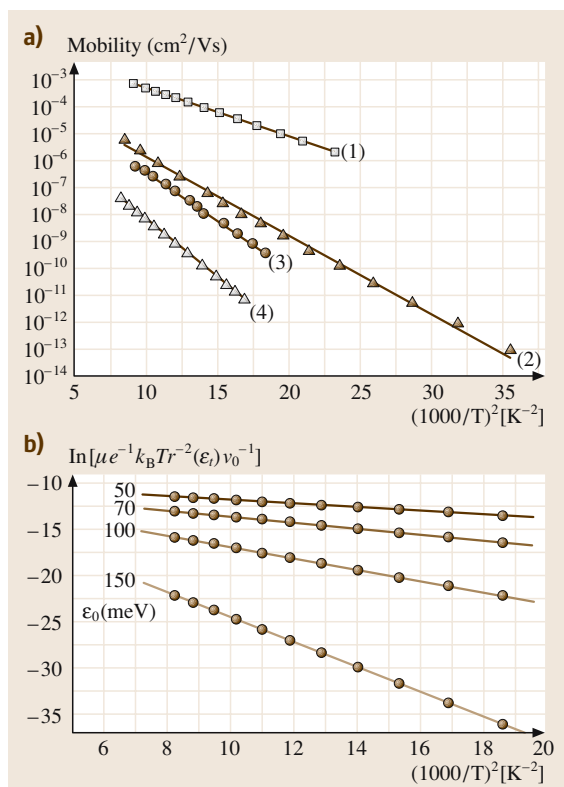


Fig. 9.14a,b Temperature dependence of the zero-field mobility in organic semiconductors. Experimental data (a): (1) di-*p*-tolylphenylamine containing (DEASP)-traps [9.69]; (2) (BD)-doped polycarbonate [9.70]; (3) (NTDI)-doped poly(styrene) [9.68]; (4) (BD)-doped TTA/polycarbonate [9.72]. Theoretical results (b) were obtained via (9.73)

with time until it approaches the thermal equilibrium value

$$\varepsilon_\infty = \frac{\int_{-\infty}^{\infty} \varepsilon \exp\left(-\frac{\varepsilon}{k_B T}\right) g(\varepsilon) d\varepsilon}{\int_{-\infty}^{\infty} \exp\left(-\frac{\varepsilon}{k_B T}\right) g(\varepsilon) d\varepsilon} = -\frac{\varepsilon_0^2}{k_B T}. \quad (9.66)$$

The time τ_{rel} required to reach this equilibrium is of key importance in the analysis of experimental data [9.65], since at $t < \tau_{\text{rel}}$ the carrier mobility decreases with time (dispersive transport) until it reaches its equilibrium, time-independent value at $t \approx \tau_{\text{rel}}$. It has been established by computer simulations that τ_{rel} strongly depends on temperature [9.18]:

$$\tau_{\text{rel}} \propto \exp\left[\left(B \frac{\varepsilon_0}{k_B T}\right)^2\right] \quad (9.67)$$

with $B \approx 1.07$. Given that the same hopping processes determine both μ and τ_{rel} , researchers were puzzled for many years by the fact that they had different coefficients B and C (in other words they have different temperature dependencies) [9.65]. Below we show how to calculate both quantities – μ and τ_{rel} – easily, and we explain their temperature dependencies (obtained experimentally and by computer simulations as expressed by (9.64), (9.65) and (9.67)).

Our theoretical approach is based on the concept of transport energy (TE), introduced in Sect. 9.3.3, where it was calculated for the exponential DOS given by (9.37). Literally repeating these calculations with the Gaussian DOS, given by (9.63), we obtain the equation [9.76, 77]

$$\exp\left(\frac{x^2}{2}\right) \left[\int_{-\infty}^{\frac{x}{\sqrt{2}}} \exp(-t^2) dt \right]^{4/3} = \left[9(2\pi)^{1/2} N_0 \alpha^3 \right]^{-1/3} \frac{k_B T}{\varepsilon_0}. \quad (9.68)$$

If we denote the solution of (9.68) as $X_t(N_0 \alpha^3, k_B T/\varepsilon_0)$, then the transport energy in the Gaussian DOS is equal to

$$\varepsilon_t = \varepsilon_0 \cdot X_t\left(N_0 \alpha^3, k_B T/\varepsilon_0\right). \quad (9.69)$$

Charge carriers perform thermally activated transitions from states with energies below the TE, ε_t , to the states with energies close to that of the TE [9.76]. Charge carriers hop downward in energy from states with energies

above the TE to the spatially nearest sites with rates determined by (9.38) and (9.39).

Now that we have clarified the relaxation kinetics of charge carriers in the Gaussian DOS, it is easy to calculate the relaxation time τ_{rel} and the drift mobility μ . We consider the case $\varepsilon_\infty < \varepsilon_t < 0$, which corresponds to all reasonable values of material parameters $N_0 \alpha^3$ and $k_B T/\varepsilon_0$ [9.76]. The energy relaxation of most carriers with energies ε in the interval $\varepsilon_\infty < \varepsilon < \varepsilon_t$ occurs via a multiple trapping-like process, well described in the literature (see, for example, *Orenstein and Kastner* [9.38] or *Marschall* [9.78]). Below ε_t the average energy of the carriers $\varepsilon(t)$ moves logarithmically downward with time t . States above $\varepsilon(t)$ achieve thermal equilibrium with states at ε_t at time t , while states below $\varepsilon(t)$ have no chance at time t to exchange carriers with states in the vicinity of ε_t . Hence the occupation of those deep states does not correspond to the equilibrium one, being determined solely by the DOS of the deep states. The system reaches thermal equilibrium when the time-dependent average energy $\varepsilon(t)$ achieves the equilibrium level ε_∞ , determined by (9.66). This happens at $t = \tau_{\text{rel}}$. Since the relaxation of carriers occurs via thermal activation to the level ε_t , the relaxation time τ_{rel} is determined by the time required for activated transitions from the equilibrium level ε_∞ to the transport energy ε_t . Hence, according to (9.40) and (9.47), τ_{rel} is determined by the expression

$$\tau_{\text{rel}} = v_0^{-1} \exp\left[\frac{2r(\varepsilon_t)}{\alpha} + \frac{\varepsilon_t - \varepsilon_\infty}{k_B T}\right]. \quad (9.70)$$

From (9.68)–(9.70) it is obvious that the activation energy of the relaxation time depends on the parameters $N_0 \alpha^3$ and $k_B T/\varepsilon_0$. Hence, generally speaking, this dependence cannot be represented by (9.67) and, if at all, the coefficient B should depend on the magnitude of the parameter $N_0 \alpha^3$. However, numerically solving (9.68)–(9.70) using the value $N_0 \alpha^3 = 0.001$, which was also used in computer simulations by *Bässler* [9.18, 65], confirms the validity of (9.67) with $B \approx 1.0$. This result is in agreement with the value $B \approx 1.07$ obtained from computer simulations [9.18, 65]. A way to describe the temperature dependence of the relaxation time τ_{rel} by (9.67) is provided by the strong temperature dependence of ε_∞ in the exponent in (9.70), while the temperature dependencies of the quantities ε_t and $r(\varepsilon_t)$ in (9.70) are weaker and they almost cancel each other out. However, if $N_0 \alpha^3 = 0.02$, the relaxation time is described by (9.67) with $B \approx 0.9$. This

shows that (9.67) can only be considered to be a good approximation.

Now we turn to the calculation of the carrier drift mobility μ . We assume that the transition time t_{tr} necessary for a carrier to travel through a sample is longer than τ_{rel} , and hence the charge transport takes place under equilibrium conditions. As described above, every second jump brings the carrier upward in energy to the vicinity of ε_t , and is then followed by a jump to the spatially nearest site with deeper energy, determined solely by the DOS. Therefore, in order to calculate the drift mobility μ , we must average the hopping transition times over energy states below ε_t , since only these states are essential to charge transport in thermal equilibrium [9.77, 80]. Hops downward in energy from the level ε_t occur exponentially faster than upward hops towards ε_t . This means that one can neglect the former in the calculation of the average time $\langle t \rangle$. The carrier drift mobility can be evaluated as

$$\mu \approx \frac{e}{k_B T} \frac{r^2(\varepsilon_t)}{\langle t \rangle}, \quad (9.71)$$

where $r(\varepsilon_t)$ is determined via (9.39), (9.63), (9.68) and (9.69). The average hopping time takes the form [9.80]

$$\langle t \rangle = \left[\int_{-\infty}^{\varepsilon_t} g(\varepsilon) d\varepsilon \right]^{-1} \times \int_{-\infty}^{\varepsilon_t} v_0^{-1} g(\varepsilon) \times \exp\left(\frac{2r(\varepsilon_t)B_c^{1/3}}{a} + \frac{\varepsilon_t - \varepsilon}{k_B T}\right) d\varepsilon, \quad (9.72)$$

where $B_c \approx 2.7$ is the percolation parameter. This numerical coefficient is introduced into (9.72) in order to warrant the existence of an infinite percolation path over the states with energies below ε_t . Using (9.63), (9.68), (9.69), (9.71) and (9.72), one obtains the following relation for the exponential terms in the expression for the carrier drift mobility:

$$\begin{aligned} \ln \left[\mu / \left(\frac{er^2(\varepsilon_t)v_0}{k_B T} \right) \right] \\ = -2 \left[\frac{4\sqrt{\pi}}{3B_c} N_0 \alpha^3 \int_{-\infty}^{X_t/\sqrt{\pi}} \exp(-t^2) dt \right]^{-1/3} \\ - \frac{X_t \varepsilon_0}{k_B T} - \frac{1}{2} \left(\frac{\varepsilon_0}{k_B T} \right)^2. \end{aligned} \quad (9.73)$$

It is (9.73) that determines the dependence of the carrier drift mobility on the parameters $N_0 \alpha^3$ and $k_B T/\varepsilon_0$.

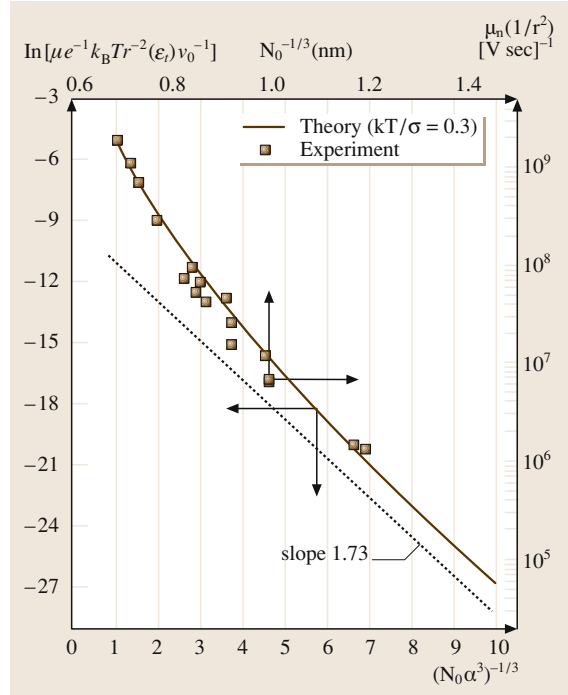


Fig. 9.15 Concentration dependence of the drift mobility evaluated from (9.73) (solid line), and the dependence observed experimentally (circles) for TNF/PE and TNF/PVK [9.79]

In Fig. 9.14b, the dependence of the drift mobility on the temperature at $N_0 \alpha^3 = 0.01$ is depicted for several values of ε_0 . The sensitivity of the mobility to temperature is clear from this picture. Comparison of these dependencies with experimental measurements of $\ln(\mu)$ versus $(1/T)^2$ [some are shown in Fig. 9.14a] provides information on the energy scale, ε_0 , of the DOS (see, for example, Bässler [9.18] and Ochse et al. [9.71]).

In Fig. 9.15, the dependence of the drift mobility on $N_0 \alpha^3$ is shown for $k_B T/\varepsilon_0 = 0.3$. Experimental data from Gill [9.81] are also shown in the figure. It is clear that the slope of the mobility exponent as a function of $(N_0 \alpha^3)^{-1/3}$ given by the theory described above agrees with the experimental data. At a very low concentration of localized states, N_0 , when the probability of carrier tunneling in space dominates the transition rate in (9.24), charge carriers hop preferentially to the nearest spatial sites. In this regime of nearest-neighbor hopping, the concentration dependence of the drift mobility is described by (9.28), as illustrated by the dashed line in Fig. 9.15.

So far we have discussed the drift mobility of charge carriers under the assumption that the concentration of charge carriers is much less than that of the localized states in the energy range relevant to hopping transport. In such a case one can assume that the carriers perform independent hopping motion and so the conductivity can be calculated as the product

$$\sigma = en\mu, \quad (9.74)$$

9.4 Concluding Remarks

Beautiful effects have been observed experimentally by studying the charge transport in disordered organic and inorganic materials. Among these, the transport coefficients in the hopping regime show enormously strong dependencies on material parameters. The dependence of the charge carrier mobility on the concentration of localized states N_0 (Fig. 9.15) spreads over many orders of magnitude, as does its dependence on the temperature T (Fig. 9.14) and on the (high) electric field strength F (Fig. 9.12). Such strong variations in physical quantities are typical, say, in astrophysics, but they are not usual in solid state physics. This makes the study of the charge transport in disordered materials absolutely fascinating. The strong dependencies of kinetic coefficients (like drift mobility, diffusivity and conductivity) in disordered materials on various material parameters makes these systems very attractive for various device applications. Since they are relatively inexpensive to manufacture too, it is then easy to understand why disordered organic and inorganic materials are of enormous interest for various technical applications.

These materials also provide a purely academic challenge with respect to their transport phenomena. While traditional kinetic theories developed for crystalline materials are largely dependent on the systems having translational symmetry, there is no such symmetry in disordered materials. However, we have shown in this chapter that it is still possible to develop a reliable theoretical approach to transport phenomena in disordered materials. Particularly interesting is the hopping transport regime. In this regime, charge carriers perform incoherent tunneling jumps between localized states distributed in space and energy. The enormously strong (exponential) dependence of the transition rates on the distances between the sites and their energies call for a completely new set of ideas compared to those for crystalline solids. Conventional transport theories based on the averaging of transition rates lead to ab-

surd results if applied to hopping transport in disordered materials. One can use ideas from percolation theory instead to adequately describe charge transport. One of the most important ideas in this field is so-called variable-range hopping (VRH) conduction. Although the rate of transitions between two localized states is a product of exponential terms that are separately dependent on the concentration of localized states N_0 , the temperature of the system T , and also on the field strength F (for high field strengths), it is generally wrong to assume that the carrier drift mobility, diffusivity or conductivity can also be represented as the product of three functions that are separately dependent on N_0 , T and F . Instead one should search for a percolation path that takes into account the exponential dependences of the hopping rates on all of these parameters simultaneously. Such a procedure, based on strong interplay between the important parameters in the exponents of the transition rates, leads to very interesting and (in some cases) unexpected results, some of which were described in this chapter. For example, it was shown that the effect of a strong electric field on transport coefficients can be accounted for by renormalizing the temperature. Most of the ideas discussed in this chapter were discussed in the early works of Mott and his coauthors (see, for example, *Mott and Davis* [9.32]). Unfortunately, these ideas are not yet known to the majority of researchers working in the field of disordered materials. Moreover, it is often believed that transport phenomena in different disordered materials need to be described using different ideas. Mott based his ideas, in particular the VRH, mostly on inorganic glassy semiconductors. Most of the researchers that are studying amorphous inorganic semiconductors (like a-Si:H) are aware of these ideas. However, new researchers that are working on more modern disordered materials, such as organic disordered solids and dye-sensitized materials, are often not aware of these very useful and powerful ideas developed by Mott and his

followers that can be used to describe charge transport in inorganic disordered systems. In this chapter we have shown that the most pronounced charge transport effects in inorganic and organic disordered materials can be successfully described in a general manner using these ideas.

Although we have presented some useful ideas for describing charge transport in disordered systems above, it is clear that the theoretical side of this field is still embryonic. There are still no reliable theories for charge transport via extended states in disordered materials. Nor are there any reliable theoretical descriptions for the spatial structure of the localized states (DOS) in organic and inorganic noncrystalline materials. All of the theoretical concepts presented in this chapter were developed using very simple models of localization centers with a given energy spectrum that are randomly distributed in

space. No correlations between the spatial positions of the sites and the energies of the electronic states at these sites were considered here. Some theoretical attempts to account for such correlations can be found in the literature, although the correlations have not been calculated ab initio: instead they are inserted into a framework of model assumptions. This shows how far the field of charge transport in disordered materials is from a desirable state. Since these materials are already widely used in various technical applications, such as field transistor manufacture, light-emitting diodes and solar cells, and since the sphere of such applications is increasing, the authors are optimistic about the future of research in this field. The study of fundamental charge transport properties in disordered materials should develop, leading us to a better understanding of the fundamental charge transport mechanisms in such systems.

References

- 9.1 A. Bunde, K. Funke, M. D. Ingram: *Solid State Ionics* **105**, 1 (1998)
- 9.2 S. D. Baranovskii, H. Cordes: *J. Chem. Phys.* **111**, 7546 (1999)
- 9.3 C. Brabec, V. Dyakonov, J. Parisi, N. S. Sariciftci: *Organic Photovoltaics: Concepts and Realization* (Springer, Berlin, Heidelberg 2003)
- 9.4 M. H. Brodsky: *Amorphous Semiconductors* (Springer, Berlin, Heidelberg 1979)
- 9.5 G. Hadziioannou, P. F. van Hutten: *Semiconducting Polymers* (Wiley, New York 2000)
- 9.6 J. D. Joannopoulos, G. Locowsky: *The Physics of Hydrogenated Amorphous Silicon I* (Springer, Berlin, Heidelberg 1984)
- 9.7 J. D. Joannopoulos, G. Locowsky: *The Physics of Hydrogenated Amorphous Silicon II* (Springer, Berlin, Heidelberg 1984)
- 9.8 A. Madan, M. P. Shaw: *The Physics and Applications of Amorphous Semiconductors* (Academic, New York 1988)
- 9.9 M. Pope, C. E. Swenberg: *Electronic Processes in Organic Crystals and Polymers* (Oxford Univ. Press, Oxford 1999)
- 9.10 J. Singh, K. Shimakawa: *Advances in Amorphous Semiconductors* (Gordon and Breach/Taylor & Francis, London 2003)
- 9.11 R. A. Street: *Hydrogenated Amorphous Silicon*, Cambridge Solid State Science Series (Cambridge Univ. Press, Cambridge 1991)
- 9.12 K. Tanaka, E. Maruyama, T. Shimada, H. Okamoto: *Amorphous Silicon* (Wiley, New York 1999)
- 9.13 J. S. Dugdale: *The Electrical Properties of Disordered Metals*, Cambridge Solid State Science Series (Cambridge Univ. Press, Cambridge 1995)
- 9.14 B. I. Shklovskii, A. L. Efros: *Electronic Properties of Doped Semiconductors* (Springer, Berlin, Heidelberg 1984)
- 9.15 I. P. Zvyagin: *Kinetic Phenomena in Disordered Semiconductors* (Moscow University Press, Moscow 1984) (in Russian)
- 9.16 H. Böttger, V. V. Bryksin: *Hopping Conduction in Solids* (Wiley, New York 1985)
- 9.17 H. Overhof, P. Thomas: *Electronic Transport in Hydrogenated Amorphous Semiconductors* (Springer, Berlin, Heidelberg 1989)
- 9.18 H. Bässler: *Phys. Status Solidi B* **175**, 15 (1993)
- 9.19 P. W. Anderson: *Phys. Rev.* **109**, 1492 (1958)
- 9.20 A. L. Efros, M. E. Raikh: Effects of Composition Disorder on the Electronic Properties of Semiconducting Mixed Crystals. In: *Optical Properties of Mixed Crystals*, ed. by R. J. Elliott, I. P. Ipatova (Elsevier, New York 1988)
- 9.21 D. Chattopadhyay, B. R. Nag: *Phys. Rev. B* **12**, 5676 (1975)
- 9.22 J. W. Harrison, J. R. Hauser: *Phys. Rev. B* **13**, 5347 (1976)
- 9.23 I. S. Shlimak, A. L. Efros, I. V. Yanchev: *Sov. Phys. Semicond.* **11**, 149 (1977)
- 9.24 S. D. Baranovskii, A. L. Efros: *Sov. Phys. Semicond.* **12**, 1328 (1978)
- 9.25 P. K. Basu, K. Bhattacharyya: *J. Appl. Phys.* **59**, 992 (1986)
- 9.26 S. Fahy, E. P. O'Reily: *Appl. Phys. Lett.* **83**, 3731 (2003)
- 9.27 V. Venkataraman, C. W. Liu, J. C. Sturm: *Appl. Phys. Lett.* **63**, 2795 (1993)
- 9.28 C. Michel, P. J. Klar, S. D. Baranovskii, P. Thomas: *Phys. Rev. B* **69**, 165211–1 (2004)
- 9.29 T. Holstein: *Philos. Mag. B* **37**, 49 (1978)

- 9.30 H. Scher, T. Holstein: *Philos. Mag.* **44**, 343 (1981)
- 9.31 A. Miller, E. Abrahams: *Phys. Rev.* **120**, 745 (1960)
- 9.32 N. F. Mott, E. A. Davis: *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford 1971)
- 9.33 A. L. Efros, B. I. Shklovskii: *J. Phys. C* **8**, L49 (1975)
- 9.34 M. Pollak: *Disc. Faraday Soc.* **50**, 13 (1970)
- 9.35 S. D. Baranovskii, A. L. Efros, B. L. Gelmont, B. I. Shklovskii: *J. Phys. C* **12**, 1023 (1979)
- 9.36 I. Shlimak, M. Kaveh, R. Ussyshkin, V. Ginodman, S. D. Baranovskii, H. Vaupel, P. Thomas, R. W. van der Heijden: *Phys. Rev. Lett.* **75**, 4764 (1995)
- 9.37 S. D. Baranovskii, P. Thomas, G. J. Adriaenssens: *J. Non-Cryst. Solids* **190**, 283 (1995)
- 9.38 J. Orenstein, M. A. Kastner: *Solid State Commun.* **40**, 85 (1981)
- 9.39 M. Grünewald, P. Thomas: *Phys. Status Solidi B* **94**, 125 (1979)
- 9.40 F. R. Shapiro, D. Adler: *J. Non-Cryst. Solids* **74**, 189 (1985)
- 9.41 D. Monroe: *Phys. Rev. Lett.* **54**, 146 (1985)
- 9.42 B. I. Shklovskii, E. I. Levin, H. Fritzsche, S. D. Baranovskii: Hopping photoconductivity in amorphous semiconductors: dependence on temperature, electric field and frequency. In: *Advances in Disordered Semiconductors*, Vol. 3, ed. by H. Fritzsche (World Scientific, Singapore 1990) p. 3161
- 9.43 S. D. Baranovskii, F. Hensel, K. Ruckes, P. Thomas, G. J. Adriaenssens: *J. Non-Cryst. Solids* **190**, 117 (1995)
- 9.44 M. Hoheisel, R. Carius, W. Fuhs: *J. Non-Cryst. Solids* **63**, 313 (1984)
- 9.45 P. Stradins, H. Fritzsche: *Philos. Mag.* **69**, 121 (1994)
- 9.46 J.-H. Zhou, S. D. Baranovskii, S. Yamasaki, K. Ikuta, K. Tanaka, M. Kondo, A. Matsuda, P. Thomas: *Phys. Status Solidi B* **205**, 147 (1998)
- 9.47 B. I. Shklovskii, H. Fritzsche, S. D. Baranovskii: *Phys. Rev. Lett.* **62**, 2989 (1989)
- 9.48 S. D. Baranovskii, T. Faber, F. Hensel, P. Thomas, G. J. Adriaenssens: *J. Non-Cryst. Solids* **198-200**, 214 (1996)
- 9.49 R. Stachowitz, W. Fuhs, K. Jahn: *Philos. Mag. B* **62**, 5 (1990)
- 9.50 S. D. Baranovskii, T. Faber, F. Hensel, P. Thomas: *Phys. Status Solidi B* **205**, 87 (1998)
- 9.51 S. D. Baranovskii, T. Faber, F. Hensel, P. Thomas: *J. Non-Cryst. Solids* **227-230**, 158 (1998)
- 9.52 A. Nagy, M. Hundhausen, L. Ley, G. Brunst, E. Holzenkämpfer: *J. Non-Cryst. Solids* **164-166**, 529 (1993)
- 9.53 C. E. Nebel, R. A. Street, N. M. Johanson, C. C. Tsai: *Phys. Rev. B* **46**, 6803 (1992)
- 9.54 H. Antoniadis, E. A. Schiff: *Phys. Rev. B* **43**, 13957 (1991)
- 9.55 K. Murayama, H. Oheda, S. Yamasaki, A. Matsuda: *Solid State Commun.* **81**, 887 (1992)
- 9.56 C. E. Nebel, R. A. Street, N. M. Johanson, J. Kocka: *Phys. Rev. B* **46**, 6789 (1992)
- 9.57 B. I. Shklovskii: *Sov. Phys. Semicond.* **6**, 1964 (1973)
- 9.58 M. Grünewald, B. Movaghar: *J. Phys. Condens. Mat.* **1**, 2521 (1989)
- 9.59 S. D. Baranovskii, B. Cleve, R. Hess, P. Thomas: *J. Non-Cryst. Solids* **164-166**, 437 (1993)
- 9.60 S. Marianer, B. I. Shklovskii: *Phys. Rev. B* **46**, 13100 (1992)
- 9.61 B. Cleve, B. Hartenstein, S. D. Baranovskii, M. Scheidler, P. Thomas, H. Baessler: *Phys. Rev. B* **51**, 16705 (1995)
- 9.62 M. Abkowitz, M. Stolka, M. Morgan: *J. Appl. Phys.* **52**, 3453 (1981)
- 9.63 W. D. Gill: *J. Appl. Phys.* **43**, 5033 (1972)
- 9.64 S. J. Santos Lemus, J. Hirsch: *Philos. Mag. B* **53**, 25 (1986)
- 9.65 H. Bässler: *Advances in Disordered Semiconductors*. In: *Hopping and Related Phenomena*, Vol. 2, ed. by M. Pollak, H. Fritzsche (World Scientific, Singapore 1990) p. 491
- 9.66 G. Schönherr, H. Bässler, M. Silver: *Philos. Mag. B* **44**, 369 (1981)
- 9.67 P. M. Borsenberger, H. Bässler: *J. Chem. Phys.* **95**, 5327 (1991)
- 9.68 P. M. Borsenberger, W. T. Gruenbaum, E. H. Magin, S. A. Visser: *Phys. Status Solidi A* **166**, 835 (1998)
- 9.69 P. M. Borsenberger, W. T. Gruenbaum, E. H. Magin, S. A. Visser, D. E. Schildkraut: *J. Polym. Sci. Polym. Phys.* **37**, 349 (1999)
- 9.70 A. Nemeth-Buhin, C. Juhasz: Hole transport in 1,1-bis(4-diethylaminophenyl)-4,4-diphenyl-1,3-butadiene. In: *Hopping and Related Phenomena*, ed. by O. Millo, Z. Ovadyahu (Racah Institute of Physics, The Hebrew University Jerusalem, Jerusalem 1995) pp. 410-415
- 9.71 A. Ochse, A. Kettner, J. Kopitzke, J.-H. Wendorff, H. Bässler: *Chem. Phys.* **1**, 1757 (1999)
- 9.72 J. Veres, C. Juhasz: *Philos. Mag. B* **75**, 377 (1997)
- 9.73 U. Wolf, H. Bässler, P. M. Borsenberger, W. T. Gruenbaum: *Chem. Phys.* **222**, 259 (1997)
- 9.74 M. Grünewald, B. Pohlmann, B. Movaghar, D. Würtz: *Philos. Mag. B* **49**, 341 (1984)
- 9.75 B. Movaghar, M. Grünewald, B. Ries, H. Bässler, D. Würtz: *Phys. Rev. B* **33**, 5545 (1986)
- 9.76 S. D. Baranovskii, T. Faber, F. Hensel, P. Thomas: *J. Phys. C* **9**, 2699 (1997)
- 9.77 S. D. Baranovskii, H. Cordes, F. Hensel, G. Leising: *Phys. Rev. B* **62**, 7934 (2000)
- 9.78 J. M. Marshall: *Rep. Prog. Phys.* **46**, 1235 (1983)
- 9.79 W. D. Gill: *J. Appl. Phys.* **43**, 5033 (1972)
- 9.80 O. Rubel, S. D. Baranovskii, P. Thomas, S. Yamasaki: *Phys. Rev. B* **69**, 014206-1 (2004)
- 9.81 W. D. Gill: Electron mobilities in disordered and crystalline tritrofluorenone. In: *Proc. Fifth Int. Conf. of Amorphous and Liquid Semiconductors*, ed. by J. Stuke, W. Brenig (Taylor and Francis, London 1974) p. 901
- 9.82 S. D. Baranovskii, I. P. Zvyagin, H. Cordes, S. Yamasaki, P. Thomas: *Phys. Status Solidi B* **230**, 281 (2002)

Dielectric Res

10. Dielectric Response

Nearly all materials are dielectrics, and the measurement of their dielectric response is a very common technique for their characterisation. This chapter is intended to guide scientists and engineers through the subject to the point where they can interpret their data in terms of the microscopic and atomistic dynamics responsible for the dielectric response, and hence derive useful information appropriate to their particular needs. The focus is on the physical concepts underlying the observed behaviour and is developed from material understandable by an undergraduate student. Emphasis is placed on the information content in the data, and the limits to be placed on its interpretation are clearly identified.

Generic forms of behaviour are identified using examples typical of different classes of material, rather than an exhaustive review of the literature. Limited-range charge transport is included as a special item. The theoretical concepts are developed from a basic level up to the ideas current in the field, and the points where these are controversial have been noted so that the

10.1	Definition of Dielectric Response	188
10.1.1	Relationship to Capacitance	188
10.1.2	Frequency-Dependent Susceptibility	188
10.1.3	Relationship to Refractive Index	189
10.2	Frequency-Dependent Linear Responses	190
10.2.1	Resonance Response	190
10.2.2	Relaxation Response	192
10.3	Information Contained in the Relaxation Response	196
10.3.1	The Dielectric Increment for a Linear Response χ_0	196
10.3.2	The Characteristic Relaxation Time (Frequency)	199
10.3.3	The Relaxation Peak Shape	205
10.4	Charge Transport	208
10.5	A Few Final Comments	211
	References	211

readers can choose for themselves how far to rely on them.

Nearly all materials are dielectrics, that is they do not exhibit a direct-current (DC) conductivity on the macroscopic scale, but instead act as an electrical capacitance i.e. they store charge. The measurement of the dielectric response is noninvasive and has been used for material characterisation throughout most of the 20th century, and consequently a number of books already exist that cover the technique from various points of view. Those that have stood the test of time are *Debye* [10.1], *Smyth* [10.2], *McCrum et al.* [10.3], *Daniels* [10.4], *Bordewijk and Bottcher* [10.5], and *Jonscher* [10.6]. These texts cover the subject in terms of the basic physics [10.1, 5], the material properties [10.2–4], and the electrical features [10.6]. An introduction to the wide range of dielectric response measurements that are undertaken can be obtained by referring to the proceedings

publication of the International Discussion Meeting on Relaxations in Complex Systems [10.7]. In view of the enormous range of properties and materials covered by the topic it is not feasible or desirable to attempt to review the whole field in a chapter such as this. Instead the topic is approached from the viewpoint of a researcher who, having measured the dielectric *spectrum* (i.e. frequency-dependent complex permittivity) of a material sample, wishes to know what information can be taken from the measurements. Along the way the limits on the information content and the problems (and controversies) associated with the microscopic and molecular-scale interpretation will be identified. Emphasis will be placed on the physical concepts involved, but inevitably there will be some mathematical expressions whose features I aim to place in as simple a physical context as possible.

10.1 Definition of Dielectric Response

10.1.1 Relationship to Capacitance

Stated simply the dielectric response of a material is its response to an electric field within it. Let us start in the simplest way with some basic macroscopic definitions. Imagine that we have a flat slab of our material of thickness d with electrodes of area A on each opposing surface, and we apply a potential difference between opposing sides (Fig. 10.1). Since the material is a dielectric our system is a capacitor and electric charges $\pm Q$ will be stored on the surfaces between which the potential difference V is applied with Q given by

$$Q = CV, \quad (10.1)$$

where C is the capacitance of the system. For the sample geometry considered C is given by the expression

$$C = (\epsilon_0 \epsilon_r A) / d. \quad (10.2)$$

Here the factor ϵ_0 is a fundamental constant termed the permittivity of free space ($\epsilon_0 = 8.854 \times 10^{-12} \text{ F m}^{-1}$) and ϵ_r is the permittivity of the material relative to that of free space and hence is termed the relative permittivity. Of course it is a pure number, which is a material property that contains information about the way the material responds to the application of the potential difference. Since however, the replacement of our sample by a vacuum yields a capacitance given by

$$C = (\epsilon_0 A) / d, \quad (10.3)$$

i.e. $\epsilon_r = 1$, the information about the response of the material to the electric field is contained in a factor χ ,

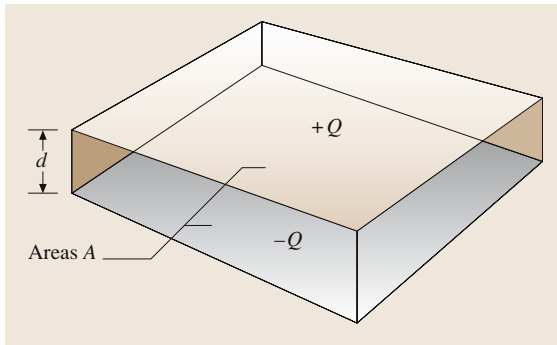


Fig. 10.1 A schematic representation of a dielectric in the form of a parallel-sided slab of surface area A and thickness d that has acquired surface charges $+Q$ and $-Q$ as a result of a potential difference applied across it. After [10.8]

which is called the susceptibility of the material and is given by

$$\chi = \epsilon_0(\epsilon_r - 1). \quad (10.4)$$

Noting that the electric field for the parallel electrode geometry described is given by $E = V/d$ and that the electrostatic flux density (electric displacement field) D within the material is $D = Q/A$, allows (10.1) and (10.2) to be transformed into the familiar constitutive relationship,

$$D = \epsilon_0 \epsilon_r E = \epsilon_0 E + \chi E \quad (10.5)$$

which is valid irrespective of the geometry. Here the response of the material has been explicitly separated out in the form of an additive term that is called the electric polarisation P , which has dimensions of charge/area or more familiarly electric dipole moment density (electric dipole moment/volume), and is given by

$$P = \chi E. \quad (10.6)$$

An electric dipole occurs when electric charges of opposite polarity are separated in space and the magnitude of the dipole moment is given by $\mu = \delta r$, where δ is the magnitude of the charge and r is the separation vector. The connection between P as defined above and the dipole moment density can be easily seen once it is recognised that the separation of the charges $\pm Q$ by the inter-electrode separation d constitutes a macroscopic dipole moment of magnitude Qd contained within the volume Ad . The definition of the polarisation as the dipole moment density makes it clear that the dielectric response to an electric field relates to the generation of a net dipole per unit volume in the material. In most cases the dielectric has a zero net dipole moment in the absence of an applied electric field and the action of the electric field is to produce one. Some materials such as electrets [10.9] and ferroelectrics lack a centre of symmetry, however, and possess a nonzero polarisation in the absence of an electric field. Therefore the definition of the polarisation P in (10.6) that is applicable to all cases is that P is the *change in net dipole moment density* produced by the applied field.

10.1.2 Frequency-Dependent Susceptibility

The outline given above has been written as if the electric field within the material was a static field ($f = 0$)

but (10.6) is equally valid if the field oscillates with a circular frequency ω ($\omega = 2\pi f$, where f is the frequency in Hertz). In this case (10.6) becomes

$$\mathbf{P}(\omega) = \chi(\omega)\mathbf{E}(\omega), \quad (10.7)$$

where $\mathbf{E}(\omega) = \mathbf{E}_0 \exp(i\omega t)$.

The fundamental reason for the dependence of $\mathbf{P}(\omega)$ upon the frequency of the alternating-current (AC) field, as in (10.7), can be envisaged by constructing a general picture of the way that a material responds to an electric field. Let us imagine that we have our material in thermal equilibrium in the absence of an electric field and we switch on a constant field at a specified time. The presence of the electric field causes the generation of a net dipole moment density (or change in one already existing). This alteration in the internal arrangement of positive and negative charges will not be instantaneous. Instead it will develop according to some equation of motion appropriate to the type of charges and dipole moments that are present. Consequently some time will be required before the system can come into equilibrium with the applied field. Formally this time will be infinity (equivalent to an AC frequency of zero), but to all intents and purposes we can regard the system as coming into equilibrium fairly rapidly after some relevant time scale, τ , with the polarisation approaching the static value $\mathbf{P} = \mathbf{P}(0)$ for $t \gg \tau$. If now we think of the electric field as reversing sign before equilibrium is reached, as is the case for an AC field at a time $t = 1/4f$ after it is switched on, it is clear that the polarisation will not have reached its equilibrium value before the field is reversed and hence that $\mathbf{P}(\omega) \lesssim \mathbf{P}(0)$, and $\chi(\omega) \lesssim \chi(0)$. The frequency dependence of the dielectric susceptibility $\chi(\omega)$ is therefore determined by the equation of motion governing the evolution of the ensemble of electric dipole moments.

In general $\chi(\omega)$ will be a complex function with a real component $\chi'(\omega)$ defining the component of $\mathbf{P}(\omega)$ that is in phase with the applied AC field $\mathbf{E}(\omega) = \text{Re}[\mathbf{E}_0 \exp(i\omega t)] = \mathbf{E}_0 \cos(\omega t)$, and $\chi''(\omega)$ defining the component that is 90° out of phase. The conventional form is given by

$$\chi(\omega) = \chi'(\omega) - i\chi''(\omega), \quad [i = \sqrt{-1}]. \quad (10.8)$$

It is easy to see that $\chi'(\omega)$ determines the net separation of charge with the dielectric in the form of a macroscopic capacitor, but the nature of $\chi''(\omega)$ is not so obvious. The answer lies in considering the *rate of change of polarisation*, $d[\mathbf{P}(\omega)]/dt$. This has the dimensions of a current density (current/area), is sometimes termed

the polarisation current density, and is given by,

$$\begin{aligned} d[\mathbf{P}(\omega)]/dt &= [\chi'(\omega) - i\chi''(\omega)] d[\mathbf{E}(\omega)]/dt \\ &= [\chi'(\omega) - i\chi''(\omega)] i\omega[\cos(\omega t) \\ &\quad + i\sin(\omega t)]\mathbf{E}_0. \end{aligned} \quad (10.9)$$

Thus $\chi''(\omega)$ determines the real component of the polarisation current density that is in phase with the electric field, i.e. $\mathbf{J}_{\text{pol}}(\omega)$ given by

$$\mathbf{J}_{\text{pol}}(\omega) = \chi''(\omega)\omega\mathbf{E}_0 \cos(\omega t) = \sigma_{\text{AC}}(\omega)\mathbf{E}_0 \cos(\omega t). \quad (10.10)$$

Here $\chi''(\omega)\omega = \sigma_{\text{AC}}(\omega)$ is the contribution to the AC conductivity due to the polarisation response to the electric field. If we remember Joule's Law for the power dissipated thermally by an electric current, i.e. power lost = IV , then we can see that $\chi''(\omega)\omega(\mathbf{E}_0)^2$ is the power dissipated per unit volume resulting from the generation of a net polarisation by the electric field, i.e. the power dissipation density. The imaginary susceptibility $\chi''(\omega)$ is often termed the power dissipation component. It arises because the electric field has to carry out work on the dielectric in order to produce a net dipole moment density. Some of this energy is stored in the charge separations and is recoverable in an equivalent way to the elastic energy stored in a spring. The rest of the energy is used to overcome the friction opposing the establishment of the net dipole density. This energy is transferred to the dielectric in an unrecoverable way, i.e. it is dissipated within the dielectric. It can be seen that $\chi''(\omega)$ is dependent upon the form of the equations of motion governing the evolution of the net dipole moment density under the action of an electric field.

10.1.3 Relationship to Refractive Index

Equation (10.7) can be regarded as relating to the polarisation response purely to an oscillating electric field, but of course all electromagnetic waves contain such a field. In general the topic of dielectric response includes the response of the material to the electric field component of an electromagnetic field, i.e. the electromagnetic spectrum of a material is a form of dielectric response. This form of response is generally characterised by a complex frequency-dependent refractive index $n^*(f)$, with

$$n^*(f) = n(f) - i\kappa(f), \quad (10.11)$$

where n is the real refractive index expressing the velocity of light in the material, v , as $v = c/n$, and κ is the

absorption coefficient, which defines the reduction of intensity of light of frequency f as it passes through the medium due to absorption of the photons by the medium, i. e. it relates to energy dissipated from the electromagnetic wave. The absorption coefficient can be determined from the Beer–Lambert law,

$$I = I_0 \exp(-4\pi\kappa fz/c). \quad (10.12)$$

For nonmagnetic materials an equivalence can be established between n^* and the relative permittivity (see [10.10] for example). If we think of two slabs of dielectric material placed in contact with an electromagnetic wave passing through them, the component of the field \mathbf{D} perpendicular to the boundary is unchanged as the wave passes from one medium to another. Essentially the same electrostatic flux passes through the same perpendicular area. Equation (10.5) gives the ratio of the electric fields in the two dielectric media to the inverse of the ratio of their relative permittivities. A comparison with Snell's law [10.10, 11] then

yields

$$\varepsilon_r = [n^*(f)]^2. \quad (10.13)$$

Measurement of the dielectric response of a material involves the determination of the polarisation and its frequency dependence in some form or other. This can be carried out in a large number of ways such as the absorption spectra of electromagnetic radiation as has been described above as well as the application of an AC electric field across a sample of defined dimensions. Those readers interested in the technical details of the measurement systems are referred to the general literature, which has an enormous number of works on these experimental techniques (see for example [10.12] for bridge techniques and [10.13] for recent microwave techniques). This chapter will have a different focus. It will in essence be a discussion of the microscopic origins of the polarisation $\mathbf{P}(\omega)$ and the physical reasons for its variation in frequency so that data obtained from such measurements can be used to gain information relevant to the nature of the material to be studied.

10.2 Frequency-Dependent Linear Responses

In this context a linear response is one in which $\mathbf{P}(\omega)$ is only dependent upon the first power of the electric field, i. e. $\chi(\omega)$ is independent of the electric field. The general form of frequency dependence expected for $\varepsilon(\omega) = [\varepsilon_0 + \chi(\omega)]$ is shown schematically in (Fig. 10.2), where it can be seen that two basic types of response can be distinguished: a resonance response at high frequencies in the quantum region, and what is termed a relaxation response at lower frequencies.

10.2.1 Resonance Response

Although our main topic in this chapter will be the relaxation behaviour I will start with the resonance response as this has been dealt with extensively [10.14] from the spectroscopic viewpoint. Here we shall approach it from the perspective of its identity as a dielectric response with the intention of identifying basic features that also occur in relaxation responses. The equation of motion is familiar: this form of response relates to a *net* electric dipole moment, ϕ , that undergoes damped simple harmonic oscillation at a natural oscillation frequency $\Omega = 2\pi\nu$ in the absence of an electric field,

$$d^2\phi/dt^2 + \gamma d\phi/dt + \Omega^2\phi = 0. \quad (10.14)$$

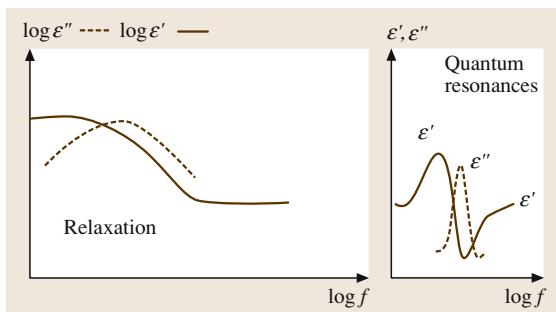


Fig. 10.2 Schematic representation of the frequency dependence of relaxation and resonance responses

The interpretation of (10.14) is not as obvious as it seems. We have to remember that we are always dealing with a sample of material that contains an enormous number of molecules or molecular moieties. Two types of situation may lead to this form of equation of motion. In one case a fluctuation in the positions of *groups* of positive and negative charges may form an oscillating net dipole moment. Plasma oscillations, found in metals [10.15], are an example of this type of behaviour. In the other case, individual molecules may possess identical dipoles oscillating independently at a frequency Ω .

Equation (10.14) then describes the behaviour of a *net* dipole moment produced by a fluctuation in the system of molecules independent of an applied electric field. Such fluctuations are deviations from the average equilibrium state produced by random impulses that act only for an infinitesimal time at $t = 0$, and (10.14) describes the subsequent evolution of the fluctuation. In both cases the damping (friction) term $\gamma d\phi/dt$, where γ is the damping factor, expresses the way in which the fluctuation dies away in the absence of an applied electric field and the net dipole moment density of the system returns to zero.

The situation applying in the case of most spectroscopic responses relates to dipole density fluctuations in which individual molecules contribute oscillating dipoles, and it is this form of behaviour that will be used to illustrate the resonance form of response. The first thing we have to do is to ask what kinds of dipole moments oscillate in our material at frequencies in the range $\nu \gtrsim 5 \times 10^{10}$ Hz where most resonance responses are found. The answer to this question lies in the electronic structure of the atoms, ions and molecules that make up our material. The uncertainty principle states the energy of a system cannot be specified at any precise instant of time. This means that the electronic structure of atoms and molecules is allowed to continuously fluctuate between quantum states as long as the average energy over a period of time remains constant. The

averaging time is determined by the uncertainty relationship between energy and time. The molecule is thus continuously moving back and forth between ground state and its excited states. These quantum fluctuations (between ground and excited states) displace the negative charge of the electron cloud with respect to the positive charge of the nucleus, and produce oscillating dipoles, termed transition dipoles (Fig. 10.3a). Their oscillation frequencies have specific values depending upon the energy difference between excited and ground quantum states of the species concerned, typically in the range $\nu \gtrsim 4 \times 10^{14}$ Hz [10.14, 16]. These oscillating dipolar fluctuations give no net contribution to the dipole moment of the molecule or atom. The uncertainty principle similarly allows the nuclei of molecules, and the molecules themselves to vibrate and fluctuate between different vibration states. In many cases these nuclear fluctuations produce dipole moments [10.10, 16, 17] (Fig. 10.3b), which typically oscillate in the frequency range 4×10^{14} Hz $\gtrsim \nu \gtrsim 5 \times 10^{10}$ Hz. As with the electronic transition dipoles, these vibration fluctuations do not contribute to the dipole moment of the molecule or system of molecules.

When an AC electric field, such as that provided by an electromagnetic wave, is applied to the material, it couples to the transition dipoles to give a force that drives the system. As a result the population of dipoles is altered such that there is now a *net* average density of transition dipole moments. This oscillates with the frequency of the *field*, i.e. the system is polarised as defined by (10.7). Since transition dipoles do not contribute to a permanent dipole moment of a molecule, the net dipole moment density produced by the electric field is referred to as an induced dipole moment density. The response of the system to an electric field oscillating at a frequency ω is essentially obtained by adding the driving force to (10.14) and determining the solution for ϕ that oscillates with the same frequency, i.e. $\phi(t) \propto e^{i\omega t}$. The components of the frequency-dependent relative susceptibility, $\chi_r(\omega) = \chi(\omega)/\epsilon_0$, then take the form

$$\chi'_r(\omega) = \frac{\chi_0 \Omega^2 (\Omega^2 - \omega^2)}{(\Omega^2 - \omega^2)^2 + \omega^2 \gamma^2}, \quad (10.15)$$

$$\chi''_r(\omega) = \frac{\chi_0 \Omega^2 \omega \gamma}{(\Omega^2 - \omega^2)^2 + \omega^2 \gamma^2}. \quad (10.16)$$

These equations exhibit the typical signature of a resonance response such as shown in Fig. 10.2. The imaginary component $\chi''_r(\omega)$ goes through a peak at $\omega = \Omega$ (the resonance frequency). The real component $\chi'_r(\omega)$ exhibits a rise in positive value as the driving

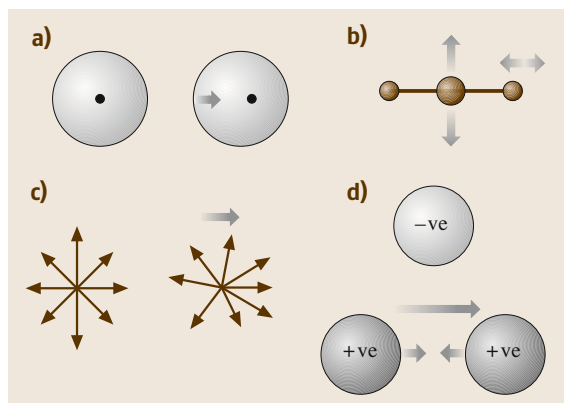


Fig. 10.3a–d Schematic representation of the various forms of dipole fluctuations: (a) atomic dipole, (b) dipoles formed by bending and stretching motion of a tri-atomic molecule of the form A–B–A, (c) net dipole fluctuation produced in a system of permanent dipoles, (d) reorienting dipole formed by transfer of an ion between two different centres of vibration. In all cases the *light arrow* shows the net dipole

frequency ω approaches the natural frequency Ω from below, passes through zero when ω and Ω are equal, and rises towards zero from a negative value when $\omega \gg \Omega$. In some cases the exact frequency dependence of the relative susceptibility can be slightly different from that given above. Typically the peak in $\chi_r''(\omega)$ is broadened due to the possibility that either the resonance frequencies of different transitions of the same molecule can be close together and their responses can overlap, or that local electrical interactions between molecules cause the transition energies of individual molecules to be slightly shifted in energy.

As long as the transition energies are sufficiently far apart to be resolved experimental data of the resonance type will yield three pieces of information, which can be related to the electronic (or vibration) structure of the molecules. These are: (a) the natural oscillation frequency Ω , (b) the damping constant γ , and (c) the amplitude factor χ_0 . The natural frequency ν is equal to the energy difference of the electronic states between which the fluctuation occurs, ξ , divided by Planck's constant i. e. $\nu = \xi/h$, and so provides information about the different quantum states in relation to one another. The amplitude factor χ_0 is proportional to the square of the transition dipole and therefore yields information on the relative rearrangement of positive and negative charges within the molecule by the transition fluctuation. Damping in these types of systems arises from the sharing of the transition energy between many energy states of the molecule and its vibrations. It removes energy from the specific oscillating dipoles for which the field produces a net dipole moment density. It may act through a delay in returning energy to the electromagnetic wave, i. e. incoherent reradiation, or by transferring it to other energy states where it cannot be reradiated. The damping therefore expresses the way that the energy transferred from the electromagnetic field to the molecule is absorbed and dissipated in the system. The damping factor γ often will have a complicated form. There is of course one other piece of information that is implied by data that fit (10.15, 16) and that is that the equation of motion for the natural oscillating dipole moments is given by (10.14). In some cases, however, the damping factor may be frequency dependent as a result of changes in the interaction between different energy states of the system that occur on the same time scale as the relaxation time, $1/\gamma$. The equation of motion will now have a different and more complicated form than that of (10.14).

The above outline of spectroscopic responses is of necessity very sketchy as it is not the main theme of this chapter, and is dealt with in detail in many stan-

dard textbooks (e.g. *Heitler* [10.14]). There are however, a number of general features that can be used as a guide to what happens in the *linear* relaxation response. In the first place the dipoles involved are a property of natural fluctuations of the system, in this case quantum fluctuations in molecules. They are not produced by the electric field. In the absence of an electric field the fluctuating dipoles do not contribute to the net dipole moment of the system, in this case individual molecules. The action of the electric field in linear responses is solely to alter the population of the fluctuations such that a net dipole moment density is produced. This is achieved in the resonance cases considered above by the production of a net density of molecules in an excited state proportional to $\chi_0 E_0$. The irreversible transfer of energy from the electric field to the system relates to the sharing of this energy between the oscillating dipoles coupled to the electric field, and many equi-energetic states of the system that do not couple directly to the electric field. The energy shared in this way is dissipated among the many connected states. Dissipation is an essential consequence of natural fluctuations in an ensemble [10.18] and expresses the requirement that the fluctuation die away to zero at long times. The function $\phi(t)$ must therefore approach zero as t tends to infinity. In the absence of an electric field dipole density fluctuations utilise energy gained transiently from the ensemble and return that energy via the dissipation mechanism. When however an electric field is present, the relative number of fluctuations with dipole moments in different directions is altered and the dissipation term irreversibly transfers energy from the electric field to the ensemble.

10.2.2 Relaxation Response

We turn now to the relaxation response. The simplest way to view this behaviour is as an overdamped oscillation of the net dipole moment density, i. e. one for which $\gamma^2 > 4\Omega^2$. There are a number of ways of addressing this situation and below I shall develop the description starting from the simplest model whose behaviour is rarely found in condensed matter.

The Debye Response

In this case we can neglect the force constant term in (10.14), i. e. the term $\Omega^2\phi$. This leads to an equation of motion with the form

$$d\phi/dt + \gamma\phi = 0. \quad (10.17)$$

The solution to this equation is the very familiar exponential form, $\phi(t) \propto e^{-t/\tau}$. Equation (10.17) can be

interpreted by taking on board the lessons from the resonance response. As before we have to view it as describing the behaviour of a natural fluctuation in our system that produces a net dipole moment density as the result of a random impulse at $t = 0$. Now however, there is no evidence for dipole oscillation, so we are not looking at the quantum fluctuations of electronic charge clouds and nuclei positions of molecules. In this case the response originates with the permanent dipoles that many molecules possess due to the asymmetry of their atomic construction. We should also remember that, though atoms do not possess a permanent dipole moment, ion pairs in a material will act as dipoles. Such systems contain a large number (ensemble) of permanent dipoles and this ensemble will obey the laws of thermodynamics. Therefore, with the exception of such materials as electrets and ferroelectrics the orientation of the permanent dipoles will be random in the absence of an electric field, i.e. the *average* net dipole moment of the system will be zero. Thermodynamic ensembles are however described by distributions that allow for fluctuations about the defined average values, thus for example canonical ensembles allow for fluctuations in energy about a defined average energy content, and grand canonical ensembles allow for fluctuations in the number of effective units (e.g. net dipole moments) as well. In the case of dipole responses we are looking at fluctuations that involve the orientations of the permanent dipoles and hence create a net dipole density (Fig. 10.3c,d). Such fluctuations are natural to the ensemble, but are transient, i.e. as in Sect. 10.2.1 $\phi(t) \rightarrow 0$ as $t \rightarrow \infty$. Equation (10.17) describes the way in which such a local fluctuation in the dipole moment density decays (regresses) to zero, i.e. the ensemble relaxes. An applied electric field couples with the permanent dipoles to produce a torque that attempts to line the dipole with the electric field vector where its energy is lowest. Consequently the linear response of the system to the application of an electric field is an increase in the population of the permanent dipole fluctuations with a component oriented in the field direction as compared to those which have components oriented in the reverse direction. This relative change in the populations of the natural fluctuations of the system gives a net dipole moment density that is driven at the frequency of the electric field as in (10.7) [10.19].

As in the resonance case the polarisation can be obtained by adding the AC driving force oscillating at frequency ω to (10.17) and determining the solution for ϕ oscillating with the same frequency. The corresponding relative susceptibility components have the

form

$$\chi_r'(\omega) = \frac{\chi_0}{1 + \omega^2/\gamma^2}, \quad (10.18)$$

$$\chi_r''(\omega) = \frac{\chi_0\omega/\gamma}{1 + \omega^2/\gamma^2}. \quad (10.19)$$

These functions show a peak in the imaginary susceptibility component, $\chi_r''(\omega)$, at a frequency $\Omega = \gamma$, which is sometimes called the loss peak frequency since $\chi_r''(\omega)$ is associated with the dissipation of energy, or equivalently the loss of energy from the driving electric field. The real component of the susceptibility, $\chi_r'(\omega)$, changes monotonically from zero at high frequencies to a limiting low-frequency value of χ_0 . This is termed the dielectric dispersion.

Equations (10.17–10.19) define what has come to be known as the Debye response after *P. Debye* who first addressed the nature of relaxation dielectric responses [10.1]. It is characterised by two pieces of information: the magnitude of the dispersion χ_0 and the damping factor γ , more usually defined via the relaxation time $\tau = 1/\gamma$ of the dipole density fluctuations. The dispersion magnitude χ_0 is a measure of the net change in dipole density fluctuations that can be produced by a unit field (i.e. $E_0 = 1 \text{ V m}^{-1}$), and is proportional to the square of the individual permanent dipole moments. As with the resonance response an exact fit between the relaxation response data and (10.18), (10.19) implies a specific form for the equation of motion of the dipole density fluctuations of the permanent dipole ensemble, i.e. that of (10.17).

Frequency-Dependent Dielectric Response in Condensed Matter

In practice the Debye response is rarely observed outside of the gas phase. Instead the experimental data can usually be characterised through fractional power laws in the frequency dependence of $\chi_r''(\omega)$ [10.6, 8, 20] in the regions away from the peak (see Fig. 10.4), i.e. for $\omega \gg \gamma$, and $\omega \ll \gamma$, giving

$$\chi_r'(\omega) \propto \chi_r''(\omega) \propto \omega^{n-1}, \quad \omega \gg \gamma, \quad (10.20)$$

$$\chi_r'(0) - \chi_r'(\omega) \propto \chi_r''(\omega) \propto \omega^m, \quad \omega \ll \gamma. \quad (10.21)$$

Here n , and m are fractional exponents, i.e. $0 < n, m < 1$. This general form was first defined empirically as the Havriliak–Negami function [10.21, 22]. A number of special cases have been identified [10.5, 8]. Thus for example the Cole–Cole function is given by $n + m = 1$. When $m = 1$, and $0 < n < 1$, the Cole–Davidson form is produced, which obeys (10.20) and

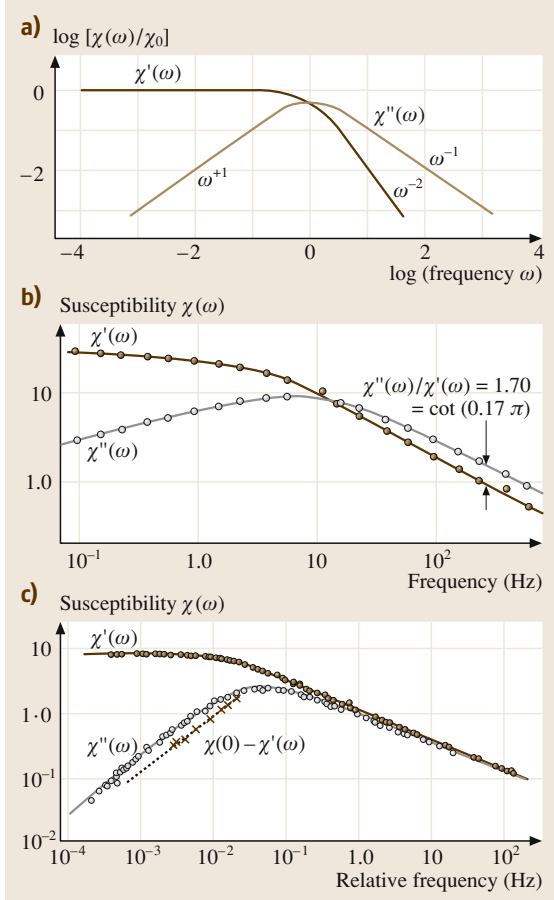


Fig.10.4 (a) Schematic drawing of the Debye response. (b), (c) Examples of measured data fitted to the response function $\chi(\omega)/\chi(0) \propto [(1 + i\omega\tau)^{n-1} {}_2F_1(1-n; 1-m; 2-n; (1+i\omega\tau)^{-1})]$ resulting from the response function $\phi(t)$ of (10.22). The function ${}_2F_1(,; ;)$ is the Gaussian hypergeometric function [10.25]. Plot (b) data from irradiated tri-glycine sulphate (TGS); Plot (c) data from polyvinylacetate. After [10.26]

follows $\chi_r''(\omega) \propto \omega^m$ ($m = 1$) for $\omega \ll \gamma$ but without $\chi_r''(\omega)$ being proportional to $\chi_r'(0) - \chi_r'(\omega)$ as in (10.21).

The physical theory of Dissado and Hill [10.23, 24] yields analytical expressions for $\chi_r'(\omega)$ and $\chi_r''(\omega)$ that contain both the Debye and Cole–Davidson functions as exact limiting cases, and the Cole–Cole and Havriliak–Negami functions as approximations with only minor differences in the curvature in the region of the peak in $\chi_r''(\omega)$. The frequency-dependent susceptibility given by this theory has come to be known as the Dissado–Hill function and is defined in terms of a hypergeometric

function [10.25], which is an infinite series with similarities to that of the exponential function. The reader is referred to [10.23, 24] for details. Unlike the empirical functions the Dissado–Hill dielectric response function has a clearly defined equation of motion for $\phi(t)$ [10.24],

$$\frac{d^2\phi}{dt^2} + \frac{[2+n+(t/\tau)]}{t} \frac{d\phi}{dt} + \frac{[n+(t/\tau)(1+m)]}{t^2} \phi = 0. \quad (10.22)$$

This equation has an analytical solution for $\phi(t)$ in terms of a confluent hypergeometric function [10.25], which has the limiting behaviour $\phi(t) \propto t^{-n}$ at times $t < \tau = 1/\gamma$ and $\phi(t) \propto t^{-(1+m)}$ when $t > \tau = 1/\gamma$. When $m = 1$ the solution is $\phi(t) \propto t^{-n} \exp(-t/\tau)$ giving the Cole–Davidson susceptibility function, and the Debye response function $\phi(t) \propto \exp(-t/\tau)$ is obtained when $n = 0$ and $m = 1$. The general analytical function for $\phi(t)$ was first derived in [10.27] and later re-derived and its equation of motion discussed in [10.23, 24] to which the reader is referred for details. The form of (10.22) illustrates clearly the development of the relaxation response from short-time high-frequency damped harmonic oscillations, characterised by (10.14), as a consequence of time-dependent damping functions and oscillation frequencies. More specifically the damping function approaches a constant value $1/\tau$ (equivalent to γ) at long times, while the oscillation frequency approaches zero. In contrast the Debye model assumes that the dipole motions are overdamped motions with no time-dependent transition from the damped oscillations occurring at short times. Of course there are many ways that time-dependent damping functions and oscillation frequencies may be introduced, most of which involve making similar approximations to that of Debye as regards the ensemble forces that control these factors, but only the specific forms of (10.22) give the power-law frequency dependencies observed in the susceptibility (10.20), (10.21). These forms are produced because the dipole density fluctuations retain some memory of the restoring and damping forces that act on it over very long periods of time, i.e. these forces are not random impulses. The reader is referred to [10.24] for more detail.

The Response Function

A different approach may be taken to deriving the frequency-dependent susceptibility of linear responses. Returning to (10.14), (10.17), and (10.20) it can be seen that the solutions for $\phi(t)$ in the absence of a driving field give the time evolution of the dipole density

fluctuation caused by a dipole generating impulse at $t = 0$, i.e. the natural motions of the system. The resulting expression for $\phi(t)$ is called the response function [10.19], and the complex susceptibility is obtained through (10.23).

$$\begin{aligned}\chi_r(\omega) &= \chi'_r(\omega) - i \chi''_r(\omega) \\ &= \text{Lt}_{\varepsilon \rightarrow 0} \left[\int_0^{\infty} \phi(t) e^{-i\omega t - \varepsilon t} dt \right].\end{aligned}\quad (10.23)$$

Here ε is an infinitesimal quantity that is taken to zero after the integral has been carried out. This equation is valid for all linear dielectric responses. The requirement that $\phi(t) \rightarrow 0$ as $t \rightarrow \infty$ for natural fluctuations ensures that $\chi'_r(\omega)$ and $\chi''_r(\omega)$ remain finite. The equations of motion, (10.14), (10.17), and (10.22) have solutions for $\phi(t)$ that can be transformed in this way to yield analytical expressions for the frequency-dependent susceptibility. An important consequence of (10.23) is that the frequency dependencies of $\chi'_r(\omega)$ and $\chi''_r(\omega)$ contain the same information since they are Laplace transforms of the same time-dependent function. This also means that a frequency-dependent conductivity $\sigma(\omega) \equiv [\omega \chi''_r(\omega)]$ contains no more information than the equivalent dielectric dispersion $\chi'_r(\omega)$. In fact the linear susceptibility components $\chi'_r(\omega)$ and $\chi''_r(\omega)$ are related to one another through the Kramers–Kronig relationships [10.5, 6]

$$\chi'_r(\omega) = \frac{2}{\pi} \int_0^{\infty} \frac{x \chi''_r(x)}{x^2 - \omega^2} dx, \quad (10.24)$$

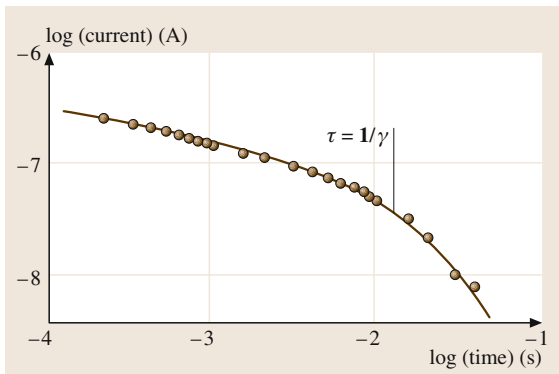


Fig. 10.5 The fit between the response function $\phi(t)/\phi(0) \propto (t/\tau)^{-n} e^{-t/\tau} {}_1F_1(1-m; 2-n; t/\tau)$ from (10.22) (where ${}_1F_1(;);$ is the confluent hypergeometric function [10.25]), and experimental data for amorphous gallium arsenide. After [10.24]

$$\chi''_r(\omega) = -\frac{2\omega}{\pi} \int_0^{\infty} \frac{\chi'_r(x)}{x^2 - \omega^2} dx. \quad (10.25)$$

The imaginary contribution from the pole at $x = \omega$ is excluded from these integrals. Although formally values of $\chi'_r(\omega)$ and $\chi''_r(\omega)$ are required over the frequency range from zero to infinity, the reciprocal relationships are adequately reproduced as long as the major part of the dispersion from any specific relaxation process is used.

Equation (10.23) means that linear dielectric responses are characterised through the time dependence of $\phi(t)$. This led *Williams and Watts* [10.28] to approach the description of the frequency dependence commonly observed by proposing that the exponential behaviour of

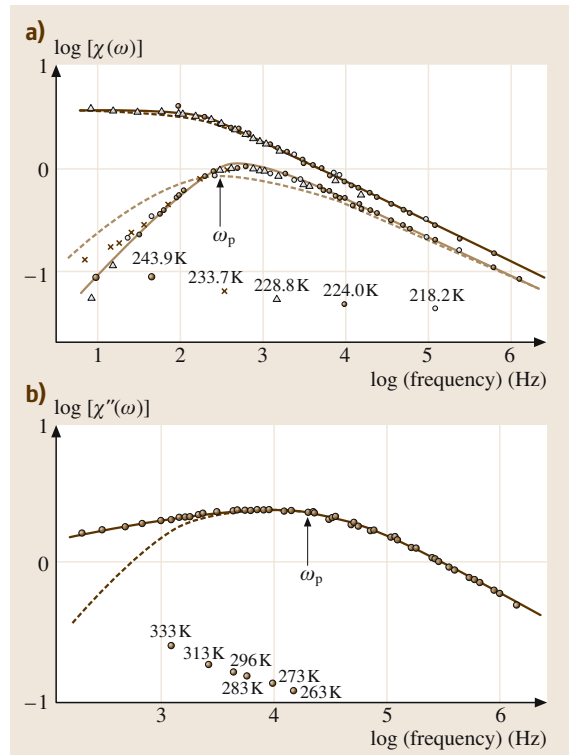


Fig. 10.6a,b Two examples of the fit between the susceptibility functions resulting from the response function $\phi(t)$ of (10.22) (continuous line) and (10.26) (broken line) and experimental master curves: (a) Nematic form of N-(2-hydroxy-4-methoxybenzylidene)-pn-butylaniline (OHMBBA), and (b) Polyallylbenzene. The shift of the representative point required to construct the master curve is marked on the plot. After [10.24]

$\phi(t)$ resulting from the Debye model, (10.17), should be generalised to the form

$$\phi(t) \propto (t/\tau)^{-n} \exp \left[-(t/\tau)^{1-n} \right]. \quad (10.26)$$

This expression is sometimes called the expanded exponential function or the Kohlrausch–Williams–Watt function, as it was later found that it was first proposed in [10.29] for mechanical responses. It is not known to possess a simple equation of motion such as (10.14), (10.17), and (10.22) but its relaxation function $\Phi(t)$, defined by $\Phi(t) = \int_t^\infty \phi(t) dt$ [10.19] obeys a relaxation equation of the form of (10.17) with a time-dependent damping factor $\gamma(t) \propto t^{-n}$. The corresponding frequency-dependent susceptibility has the same power-law form as (10.20) for $\omega \gg \gamma = 1/\tau$, but

exhibits a slowly varying decrease of slope as the frequency γ is approached from below that, with suitable choices for the value n , can approximate a power law for $\chi_r''(\omega)$ such as is defined in (10.21). The relationship between experimental data for $\phi(t)$ and that derived in the Dissado–Hill cluster model, i. e. the solution to (10.22), is shown in Fig. 10.5. It can be seen that the data and the function for $\phi(t)$ approaches zero as t tends to infinity, with the time power law $t^{-(1+m)}$, but accurate experimental data for times several decades beyond τ is required if this behaviour is to be distinguished from that of (10.26). A better means of distinguishing the two results can often be had by recourse to their appropriate frequency-dependent susceptibilities, see Fig. 10.6.

10.3 Information Contained in the Relaxation Response

As described in Sect. 10.2.2 relaxation responses contain three pieces of information. The strength of the coupling of dipole density fluctuations to the electric field characterised by χ_0 , a characteristic relaxation frequency $\gamma = 1/\tau$, where τ is the characteristic relaxation time, and the relaxation dynamics characterised by the frequency dependence of $\chi_r'(\omega)$ and $\chi_r''(\omega)$. This latter feature is open to different interpretations, as will be discussed later.

10.3.1 The Dielectric Increment for a Linear Response χ_0

The dielectric increment is proportional to the square of the permanent dipole moments that give rise to the dipole density fluctuation. It is a feature of the dielectric response that does not usually receive the most attention, mainly because a quantitative relationship to the molecular physics of the relaxation process is often difficult to achieve. Nevertheless it has been used to determine the dipole moments of polar molecules using measurement in the gas phase or if necessary dilute solutions in a non-polar solvent. In these cases the permanent molecular dipoles, μ , can be assumed to be independent of one another and to be able to adopt all orientations with equal probability in the absence of an electric field, i. e. all dipole moment orientations are at the same energy. This section starts by outlining the derivation of χ_0 for this situation even though this is not the topic area of this book and chapter. The aim is to demonstrate the procedure and bring out the assumptions involved so that the

more complicated nature of dipole density fluctuations in condensed-state materials can be better appreciated.

Independent Free Dipoles

In an electric field a dipole that is at an angle θ to the field direction is at the energy $-\mu E \cos \theta$. Those molecules aligned with the electric field are therefore at the lowest energy. The thermal motions of the molecules will however tend to randomise the dipole orientations and the probability of finding a dipole with an orientation angle θ becomes $\exp(\mu E \cos \theta / k_B T)$. The average value of $\mu \cos \theta$ is given by

$$\begin{aligned} M &= \langle \mu \cos \theta \rangle \\ &= \frac{\int_0^\pi \mu \cos \theta \exp(\mu E \cos \theta / k_B T) \sin \theta d\theta}{\int_0^\pi \exp(\mu E \cos \theta / k_B T) \sin \theta d\theta} \end{aligned} \quad (10.27)$$

and the contribution to the static polarisation is given by $N \langle \mu \cos \theta \rangle$, where N is the number of permanent dipoles per unit volume. The term independent of the electric field is zero because all orientations are equally probable in the absence of the field. Equation (10.27) results in a nonlinear function in the electric field E for M , which is called the Langevin function, $L(\mu E / k_B T)$, with,

$$L(\mu E / k_B T) = \coth(\mu E / k_B T) - (k_B T / \mu E). \quad (10.28)$$

This function saturates at unity for very high values of $\mu E/k_B T$, reflecting the total alignment of all the permanent dipoles in the electric field. At low fields defined by $\mu E/k_B T \ll 1$, $L(\mu E/k_B T)$ is a linear function of E and gives the linear dielectric increment as

$$\chi_0 = \mu^2/3k_B T. \quad (10.29)$$

Dipoles in Condensed Matter

In condensed-phase systems, particularly solids, the approximations that lead to the Langevin function and (10.29) no longer apply, and hence these expressions no longer hold. In the first place the dipoles are constrained by the local structure and in general will not be able to assume all orientations with equal probability in the absence of an electric field. In the second place we cannot assume that the dipoles are independent of one another. This dependence may arise in more than one way. For example there may be electrostatic interactions between the dipoles, such as would be responsible for the formation of ferroelectric and anti-ferroelectric states. However when the dipoles concerned are of a low concentration such as those that originate with impurities, lattice defects, interstitial ions etc., these dipole–dipole interactions may be weak. The dipoles concerned may also be arranged in such a way that even though they can adopt one or more alternative orientations their dipole–dipole interactions essentially cancel, such as might be expected in dipole glasses [10.30]. The common way to deal with this situation is to assume that a dipole representative of the average dipole in the ensemble experiences the average electric field of all the other dipoles. This is called the *mean-field* approach [10.31]. Since the mean field will be a function of the average dipole moment due to the applied electric field it is usually possible to construct an equation that can be solved to yield M and hence χ_0 . Another way in which the dipoles can interact arises because permanent dipoles are part of the lattice structure of the material. Those permanent dipoles that lead to a polarisation in the presence of an electric field must have two or more local orientations available to them, i. e. they must be able to adopt a different orientation that in the presence of an electric field has a lower energy. Any such change will inevitably alter the local atomic and molecular interactions around the dipole that has moved. This effect will travel through the structure and influence other permanent dipoles through changes in atomic and molecular positions in its environment [10.32]. The strength of such interactions will vary depending of the type of dipole and the way that it is connected to the structure. For example reorientable

dipoles formed by small interstitial (or substitution) ions may not interact very strongly with the surrounding lattice, whereas polar groups attached to a polymer chain will in many cases interact very strongly when they adopt a different orientation. Similarly the reorientation of polar molecules in liquids may be expected to distort their surrounding solvent cage and create a disturbance that will be transmitted to other polar molecules. The special feature of this form of interaction is that it is transmitted along specific directions depending upon the lattice structure and hence is nonisotropic.

Order–Disorder Ferroelectrics

These are materials in which the permanent dipoles possess two or a limited number of possible orientations. At high temperatures the dipoles are randomly distributed between the alternative orientations in the absence of an electric field. As the temperature is lowered the electrostatic field of the dipoles acts on any one dipole to make one of the orientations more preferable than the others. This causes the permanent dipole system to adopt a specific orientation at the Curie temperature T_c . The mean-field approach results in an expression for χ_0 that diverges at T_c , i. e.

$$\chi_0 \propto \mu^2/|(T - T_c)|. \quad (10.30)$$

This expression is so common to us that it is easy to overlook the physical meaning that it contains, which is much better expressed in the renormalisation group approach [10.33]. Essentially the interactions between the dipoles cause their orientation and dynamics to become correlated to some extent. As T_c is approached from above, the dipole fluctuations in the system are correlated over increasingly long distances and involve increasingly larger groups of individual dipole moments μ . The dielectric increment increases in proportion to a power of the correlation length $\xi \propto |(T - T_c)^{-\delta}|$ and a more exact form for χ_0 is

$$\chi_0 \propto \mu^2/|(T - T_c)|^\alpha. \quad (10.31)$$

At temperatures below T_c the material will possess domains in which all the dipoles are aligned together. Dipole fluctuations in this state have the opposite orientation to that of the polarity of the domain dipole that is they are *changes in net dipole moment density*, see Sect. 10.1.1. These dipole fluctuations also produce an electrostatic field that causes them to be correlated. As the temperature reduces their correlation length reduces and hence so does χ_0 . These materials show that the responding dipole in condensed-phase materials will not

always be that of individual molecular or ionic components of the system, they may in fact be groups of individual dipoles that respond as a single unit. In fact the susceptibility increment may be written in terms of an effective dipole moment that is also a power of the correlation length, i. e. $\mu_{\text{eff}} \propto (\xi)^{\alpha/2\delta}$.

Kirkwood Correlation Factor

The particular feature of the ferroelectric order–disorder transition is that the correlation between dipoles in the dipole fluctuations is temperature dependent. In general this will not be the case; instead we can expect the correlation to be dependent upon the lattice structure of the material, and hence independent of temperature except for discrete step changes when the material undergoes a phase transition. One way of allowing for these correlations is to introduce a factor g , termed the Kirkwood factor [10.10, 34], into the expression for χ_0 such that it becomes

$$\chi_0 \propto Ng\mu^2. \quad (10.32)$$

Here a value of $g = 1$ defines a system in which the dipoles are uncorrelated, i. e. they are independent of one another. Values of $g > 1$ indicate positive correlation i. e. the dipoles align in the same direction as in the ferroelectric case where g diverges as discussed. Values of $g < 1$ indicate dipoles that are anticorrelated i. e. pairs of dipoles tend to align in opposite directions. Both kinds of behaviour are known to occur.

General Features of the Temperature Dependence of the Dielectric Increment

Section 10.3.1 gives examples of behaviour in which an electric field imposes order, in terms of the polarity of the net dipole density fluctuations, upon an ensemble where the temperature acts as a disordering factor. In the case of independent free dipoles it is the applied field that attempts to align dipoles and for order–disorder ferroelectrics it is the electrostatic field of the other dipoles. It might therefore be expected that the temperature is always a factor that attempts to oppose the field and that the dielectric increment will decrease at high temperatures, but this is not always the case. For example, in situations where the dipole can adopt one of two orientations one of which is favoured by the local lattice structure, the equilibrium population will be heavily weighted towards the favoured orientation. The equilibrium value of M in such a region can be expressed as,

$$M = \mu \tanh[(\Delta_\chi/k_B T)], \quad (10.33)$$

where $2\Delta_\chi$ is the energy difference between the alternative orientations produced by the local structure. Here μ is the reorientable component of the dipole moment, see Fig. 10.7. An applied electric field that will favour the one orientation over the other will change the energy difference between alternative orientations from $2\Delta_\chi$ to $2(\Delta_\chi - \mu E)$. Differentiation with respect to E then gives the susceptibility increment as $\chi_0 \propto (\mu^2/k_B T) \cosh^{-2}(\Delta_\chi/k_B T)$, with the approximately activated form

$$\chi_0 \propto (\mu^2/k_B T) \exp(-2\Delta_\chi/k_B T) \quad (10.34)$$

holding when $\Delta_\chi/k_B T \gg 1$. Although (10.33) implies that regions exist where dipoles are aligned by the structure this does not necessarily mean that the material has a net dipole in the absence of an electric field. Such regions may be local and with dipole vectors randomly arranged by the structure. This behaviour has been found in ferroelectric ceramics [10.35, 36] in both the ferroelectric and paraelectric phases outside the transition region where T approaches T_c . Expression (10.34) indicates that in the appropriate temperature range the effect of the thermal fluctuations is effectively to increase the density of dipoles that can respond to the electric field, and that this overcomes any randomising behaviour. In contrast, at high temperatures ($\Delta_\chi/k_B T \ll 1$) the distribution of dipoles between the alternative orienta-

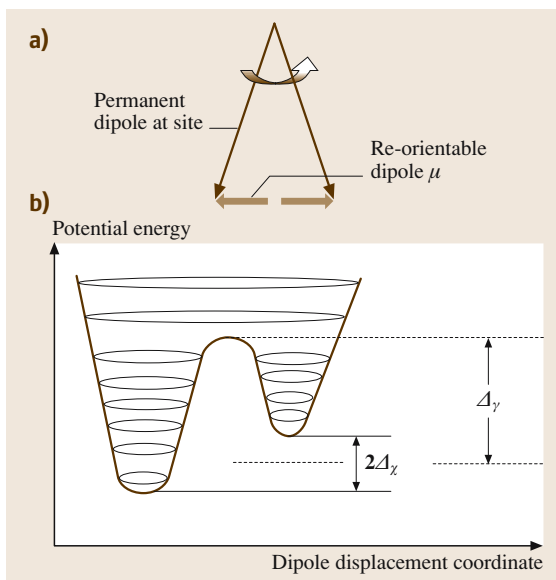


Fig. 10.7a,b Dipole reorientation between two potential wells: (a) shows the reorientable component of the dipole, (b) shows the potential-energy surface

tions becomes almost random and χ_0 approaches the free-dipole result (10.29).

In many experimental situations the value of the dielectric increment is essentially independent of temperature. It is difficult to see how this can occur in an ensemble where the dipole density fluctuations are produced by fluctuations in thermal energy about the average value, which couple to the electric field via changes in the heat content as in (10.27). However it may be possible to conceive of this behaviour as due to fluctuations in the configuration entropy of the molecular system, of which the dipole is a part, that take place without any change in the heat content. The effect of the field would be to change the configuration entropy S rather than the heat content H . As a result the susceptibility would be independent of temperature. This picture implies that we must think of the dipoles in this case not as local elements embedded within the material matrix moving in a fixed local potential, but as an integral part of the matrix whose dynamics is described by fluctuations in the Gibbs free energy $G = H - TS$ of the whole ensemble. In this case correlations between dipoles would be expected to occur mainly via the indirect route through their interaction with their local environment, rather than their direct electrostatic interactions.

Equation (10.31) describes the behaviour of a system undergoing an order–disorder transition *among the permanent dipole orientations*. A similar behaviour will be found for the relaxation response of a first-order ferroelectric or dipole alignment transition [10.37]. In general phase transitions will not give rise to a divergence in χ_0 , which occurs because the phase transition in these cases is defined through the dipole, i. e. the dipole orientation is the order parameter. In other types of phase transition the dipoles are not the primary cause and what can be expected is an abrupt change of χ_0 as the dipoles find themselves embedded into a different lattice structure with different local potentials and orientation positions, different ensemble energies, and different correlations with one another and the material matrix.

The Information Content of the Dielectric Increment

As is clear from the above discussion it is not easy to make definite quantitative statements about the dipole system based on the dielectric increment. The basic reason for this situation is that the measurements are made on a macroscopic sample that contains an ensemble of an enormous number of dipoles, up to $\approx 10^{28} \text{ m}^{-3}$. The description of such systems can be carried through if the

elements, here electric dipoles, are independent and their orientation is defined by a static local potential; as discussed above this will not be the case in general. More typically the dipoles will be correlated with the matrix in which they are embedded and/or one another. This means that the dipoles that are involved in the dipole density fluctuations are not site dipoles but groups of molecules/ions including dipoles, i. e. the responding features have a size intermediate between that of the molecule/unit cell and that of the sample. Determination of the temperature dependence will give some clues as to how to regard the dipole system through the definition of an effective dipole. The way the effective dipole changes with temperature will allow some interpretation of the kind of system that is present. Variation with other control parameters will produce more information, and systematic variation of the structure, for example replacement of side groups in polymers by longer or different side groups, or substitution of impurity ions by similar ions of different oxidation state or ionic radius, will help to identify the local dipole moment contributing the dielectric increment. However, even if the form of the site dipole is known, the quantitative evaluation of a factor such as the Kirkwood factor g (10.32) is not trivial. In condensed matter, as can be seen from Fig. 10.7, only a component of the site dipole is likely to be involved in reorientation. Unless the local structure is very well known it will not be possible to determine the actual value of the reorientable component in order to obtain a quantitative estimate of g . What can be achieved is a fingerprint of the dipole fluctuations in the material that can be used to characterise it. However at best this will be a partial picture of the dipole fluctuations in the material and information gained from the relaxation time and the relaxation dynamics should be used to enhance it further. In this way a holistic view of the dipole fluctuation can be attempted. It is important to realise that the picture obtained from these three features must be complementary. It is not acceptable to regard them as three independent features, as in fact they just yield different facets of the same process.

10.3.2 The Characteristic Relaxation Time (Frequency)

Equation (10.14), (10.17), (10.22), and (10.23) define a characteristic relaxation rate γ or relaxation time $\tau = 1/\gamma$ for the dipole density fluctuations. In the case of the Debye response, whose susceptibility functions are given by (10.18) and (10.19), γ is the frequency at which the imaginary (dielectric loss) component $\chi''(\omega)$

exhibits a peak. It has therefore become customary to determine the dependence of the relaxation time (rate) upon the control parameters (e.g. temperature, pressure, etc.) that are varied via that of the loss peak frequency. As long as the frequency dependence of $\chi_r''(\omega)$ (i.e. the loss peak shape) remains unchanged this procedure is valid because essentially the dielectric response investigated can be treated as a single composite process, even though it has a wider frequency dependence than that predicted for the independent free dipoles. Both the theoretical equation of motion (10.22) and response function (10.26) do in fact describe the response as a single composite process with a characteristic relaxation rate (time). However the frequency of the peak in $\chi_r''(\omega)$ (loss peak frequency) predicted from (10.22) is not γ but γ multiplied by a numerical factor depending upon the power-law exponents, n and m . In many cases the loss peak is very broad ($n \rightarrow 1, m \rightarrow 0$) and it is difficult to locate the peak precisely and to be sure that the point located is at the same position with respect to the functional dependence of $\chi_r''(\omega)$ upon ω . Under these circumstances a better procedure is to construct a master curve, which is done by plotting $\chi_r''(\omega)$ as a function of ω in log-log coordinates. Translation of the data along the $\log(\omega)$ and $\log[\chi_r''(\omega)]$ axes will bring the data into coincidence if the susceptibility frequency dependence is unchanged. The translation required to achieve coincidence gives the dependence of the susceptibility increment [$\log(\chi'')$ -axis] and characteristic relaxation frequency [$\log(\omega)$ -axis] on the controlled variable. For example it gives the ratios $\gamma(T_1)/\gamma(T_2)$ [or $\chi_0(T_1)/\chi_0(T_2)$] for the temperature change T_1 to T_2 . This technique also has the advantage of illustrating clearly whether or not the frequency dependence is independent of the variation in temperature (or other parameter), i.e. whether or not the different sets of data can be brought into coincidence. It can also be used to determine any relationship between χ_0 and γ . This is done by selecting a reference point (e.g. the point $\chi_0 = A, \omega = B$) and marking the position on the master curve of this point from each data set after it has been translated to achieve coincidence. A trace is formed giving the dependence of $\log[(\chi_0)^{-1}]$ as a function of $\log(\gamma^{-1})$.

The relaxation rate is the dielectric response feature that shows most dependence upon the variation in the control parameters and so is the feature that is most often studied. In the following sections I will outline some of the most common types of behaviour and discuss their implications for the physics of the relaxation process.

Site Dipole Reorientation

The simplest form of relaxation rate is that described by Debye for independent molecular dipoles suspended in a viscous continuum. As described in Sect. 10.3.1 these dipoles are regarded as free to adopt any orientation in the absence of an electric field. Relaxation of a dipole density fluctuation involves the rotation of the molecular dipoles in the fluctuation to a state in which the net dipole density is zero. In such a situation the rotation of each individual dipole occurs at the same speed determined by the viscosity, η , of the medium, and the relaxation time ($\tau = 1/\gamma$) of the dipole density fluctuation is governed by that speed. For a molecular dipole whose effective length is a the relaxation time has the form [10.10],

$$\tau \propto \eta a^3 / k_B T. \quad (10.35)$$

The more viscous the medium, or the bigger the molecular dipole, the slower it rotates and the slower will be the relaxation of a fluctuation, giving a net dipole moment to the ensemble.

Of course the conditions for this behaviour to be exactly applicable cannot be met except in a gaseous medium. Condensed-phase materials are not continua. Even liquids possess a local structure and molecular dipoles will either be part of that structure if they are contributed by the medium, or will be surrounded by a solvent shell if they are dissolved in the medium. In solids the molecular (or ionic) dipole is of necessity part of the structural matrix, and even though this must be irregular enough to permit rotational displacement to at least one other orientation the matrix can be expected to maintain some structural correlation to distances well away from the site of a reorientable dipole. These are the conditions that must be included in any description of the relaxation frequency (time).

The first point of departure from the picture of a free dipole in a continuum is that the dipole will possess only a limited number of orientations that it can adopt. Consequently there will be a potential barrier between these alternative equilibrium orientations. The rate of transition between alternative orientations and hence the relaxation frequency will be determined by the rate at which a dipole or, to be more specific, the atoms or ions that form the local dipole can pass over the potential barrier to switch orientation, as shown for example in Fig. 10.7. In this case the relaxation frequency will possess an activated (Arrhenius) form where the activation energy Δ_γ is the mean potential barrier height between the alternative orientations, i.e.

$$\gamma = 1/\tau = A \exp(-\Delta_\gamma/k_B T). \quad (10.36)$$

The expression to be used for the pre-exponential factor A depends on the way in which the atoms/ions comprising the local dipole pass through the transition region at the top of the barrier [10.38, 39]. In the schematic drawing of Fig. 10.7 the dipole is described as having an atom (ion) at its head that performs quantum oscillations in one of two potential wells. As long as it can be assumed that there is a thermal equilibrium between all the vibration states and that the dipole head passes into the alternative well in a single transit of the barrier region, then $A = \nu$, where ν is the frequency of the quantum vibrations at the bottom of the wells. This result continues to hold even if thermal equilibrium is established only for the states at the bottom of the well as long as the effective friction acting on the dipole head in the barrier region is weak. The type of potential surface with these properties is one that remains essentially rigid during the actual transit of the barrier region, which takes place in a time typically of $\approx 10^{-14}$ s. The activated factor in (10.36) expresses the thermal probability of finding a dipole in a quantum state at the top of the barrier. The other extreme situation occurs when the friction ζ_d in the barrier region is high, for which $A \propto (1/\zeta_d)$. This occurs when the transit of the barrier region is slow enough to allow interactions with neighbouring vibrating atoms to overdamp the motion. In this case the potential surface distorts during the transit of the barrier. Such situations can be expected when the barrier is broad and ill-defined, and correspond to local structures that are flexible, such as may be expected in viscoelastic materials. A general expression $A = \lambda_r(\nu/\nu_b)$ has been developed by *Grote and Hynes* [10.39] where λ_r is a function that describes the change from low to high friction and ν_b is the quantum frequency in the barrier region.

An interesting consequence of this type of potential surface is that, regardless of the magnitude of the barrier energy Δ_γ , a temperature should exist below which reorientation over the barrier would take so long that any dipole fluctuation would essentially remain unrelaxed, i. e. the dipole system becomes frozen. However when the moving atom in the permanent dipole is a hydrogen atom this is not the case; relaxation can occur by the tunnelling of the hydrogen atom through the barrier [10.40]. This has been demonstrated by experiments on deuterated oxidised polyethylene molecules at millikelvin temperatures [10.41, 42]. In this case the relaxation frequency is determined by the tunnelling probability of the deuterium/hydrogen atom through the barrier, which is dependent upon the atomic mass, the barrier width and height, but *not the temperature*, i. e. the

relaxation frequency becomes temperature independent at temperatures below ≈ 100 mK.

Relaxation on a Free Energy Surface

The situation described in the previous section is one in which the dipole moves on a potential surface provided by the surrounding structural matrix. The only dynamic effect of the matrix is via elastic and inelastic interaction between the quantum vibrations of the dipole and the matrix. In many cases however, the atoms (ions) comprising the dipole will cause the displacement of the centres of motion of the surrounding atoms during its transit between alternative orientations. In this case the expression for the relaxation frequency has to refer to the group of atoms affected as a unit, and an appropriate form is that derived by *Eyring* [10.43] for chemical reactions

$$\gamma = (k_B T/h) \exp(-G^\# / k_B T). \quad (10.37)$$

Here $G^\#$ is the change in Gibbs free energy on passing from the ground state to the transition state in the process of reorientation. The barrier is now a free energy rather than a potential barrier and reflects the need for the involvement of displacements in a number of atoms, ions or molecules in order to achieve the dipole reorientation. If we refer again to Fig. 10.7 the difference is that the normal coordinate of the quantum vibrations in the barrier region is a mixture of several different normal coordinates of the surrounding matrix as well as that of the dipole in the well. In general $G^\#$ will be composed of an activation entropy contribution $S^\#$ as well as an activation enthalpy contribution $H^\#$ with $G^\# = H^\# - TS^\#$, and both will be properties of the group of atoms/molecules involved and their structural relationship. The expression for γ therefore takes a form similar to that of (10.36):

$$\begin{aligned} \gamma &= \left[(k_B T/h) \exp(S^\# / k) \right] \exp(-H^\# / k_B T) \\ &= A_{\text{eff}} \exp(-H^\# / k_B T). \end{aligned} \quad (10.38)$$

The activation entropy $S^\#$ will reflect the configuration rearrangement required for the dipole to reorient. Thus for example, when reorientation requires the surrounding matrix to adopt a more irregular (disordered) arrangement so as to remove a steric hindrance to reorientation the transition region entropy will be greater than that of the dipole in the bottom of the well and $S^\#$ will be positive. Alternatively the transition region may require specific local arrangements in order that the dipole can avoid such hindrances. In this case the entropy of the transition state will be less than that of the

dipole in the bottom of the well and $S^\#$ will be negative. A nonzero value of $S^\#$ will therefore lead to a relaxation frequency with an effective pre-exponential factor, A_{eff} , that is either greater or less than $k_B T/h$.

An interesting variation on this behaviour has been suggested by Hill and Dissado [10.44] who showed that several experimental relaxation frequencies could be described via what they termed an activated tunnelling expression. Here they allowed the possibility that the transfer between alternative orientations could take place by thermal activation to an energy state for which the tunnelling of hydrogen atoms through the barrier was feasible. An expression for the relaxation frequency was obtained by determining the optimum transfer rate, with all processes considered from that of an unactivated quantum tunnelling to activation over the top of the barrier. If the energy difference between the alternative orientations is zero the resulting relaxation frequency had the form

$$\gamma \propto \exp(BT) \quad (10.39)$$

over a considerable temperature range, where B is a temperature-independent factor dependent upon the tunnelling distance, the barrier height and shape, and the mass of the tunnelling particle. If the *barrier* in this work is taken to be a free-energy barrier rather than a potential barrier, the analysis can be seen to be equivalent to a situation whereby the dipole relaxes by finding a route between the alternative orientations that allows for lower values of $H^\#$ at the expense of ordering the surrounding molecules/atoms, i. e. a negative $S^\#$. Routes with higher values of $H^\#$ require values of $S^\#$ that are less negative. In this case the optimum relaxation frequency will take a form such as (10.39) and only at very high temperatures will γ become purely activated with $S^\# \approx 0$. The picture of dielectric relaxation provided by this interpretation is consistent with the *defect diffusion* mechanism [10.45], in which defects diffusing in the structural matrix, such as a kink in a polymer chain or a dislocation in a crystal, lower or remove barriers when they reach a dipole, allowing it to take an alternative orientation of equal energy. When the defect moves on, the dipole is locked into the new position until another defect arrives.

The Glass Transition

A glass is essentially a material that has become macroscopically rigid without attaining its thermodynamically favoured crystalline state. The manner in which this occurs for a liquid or viscoelastic (rubbery) material that has been rapidly cooled has been the subject of in-

tense investigation over a large number of years (see Angell [10.46]). A simple, some would say oversimple, view of the situation is to regard it as a concatenation of two effects. In the first place supercooling the liquid phase will at some temperature result in a situation where the liquid no longer possesses a heat content in excess of the crystal state at that temperature [10.47]. The liquid will be unstable with respect to a disordered solid possessing only local crystalline order. In the second place the lowering of the temperature during the supercooling will cause the lattice to contract thereby introducing and increasing barriers to local molecular/atomic motions. At some temperature the thermal fluctuations [described by the Boltzman factor in (10.36), (10.37)] that are responsible for raising the site energy to that of the barrier become so rare that the time required for rearrangement becomes enormously long. At this temperature, termed T_g , the structure is essentially locked into a macroscopically rigid state, termed a glass. The glass transition cannot be regarded as a thermodynamic transition of state (unlike a melting/crystallisation temperature for example) and there are different ways of defining and determining T_g (see [10.46]).

During the approach to the glass state the rate of reorientation of permanent dipoles will become slower and eventually reach zero as shown in Fig. 10.8. It is common to denote the response due to a set of dipoles that are frozen during glass formation as the α -response. A second response can be seen in the figure: termed the β -response. This is provided by dipoles that are able to reorient without requiring any substantial rearrange-

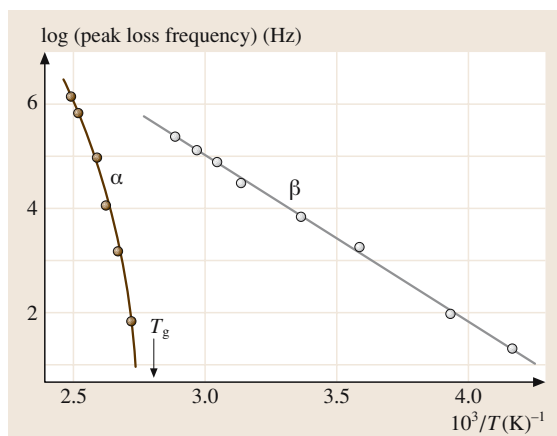


Fig. 10.8 The Arrhenius plot of the glass-forming polyvinylchloride system showing the slowing down of the α -response ($T > T_g$) as T_g is approached and the activated β -response. After [10.8]

ment of the surrounding structural matrix. It should not be expected however that this response is due just to a dipole reorientation with respect to the molecule it is attached to. In many cases the β -response involves the displacement of the molecule or part thereof as a whole [10.48]. In polymers this is a local *inter-chain* motion and either the free-energy expression (10.38) or the potential-energy expression (10.36) will apply, depending on whether the surrounding chains remain rigid during the relaxation or rearrange locally. These dipoles are active in the glass state and can be expected to have a relaxation rate of the form of (10.36), i. e. reorientation over a potential barrier. In the case of the α -response it is clear that the relaxations must involve displacements in a number of molecules/atoms other than just those comprising the permanent dipole, and hence it is instructive to discuss the behaviour in terms of the rate expression (10.37). What can be seen is that, as the temperature at which the system becomes rigid is approached, the gradient in the Arrhenius plot gets steeper, and $H^\#$ therefore becomes larger. The non-thermally-activated pre-exponential factor, A_{eff} , in (10.36) is greater than $k_B T/h$ and hence $S^\# > 0$. As the temperature approaches T_g there is an increase in A_{eff} , by many decades in frequency, which must be due to an increase in $S^\#$. These changes in $H^\#$ and $S^\#$ indicate that, as T_g is approached, dipole-orientation relaxation not only requires an increased amount of energy in order to enter the transition (barrier) region but also a larger amount of configuration disordering in the surrounding structure that makes up the molecular/atom group involved in relaxation. Although it is possible that such a situation may come about because reducing the temperature produces a local increase in density that increases steric hindrances for the same group of atoms and molecules, it is more likely that the number of molecules that are displaced in order to allow the dipole to pass through the transition region has increased. These considerations are consistent with a structure that is becoming either tangled or interlocked as the temperature decreases. Detailed expressions based on these concepts but involving macroscopic parameters have been attempted (see for example [10.46, 49–52]).

The glass formation discussed above has a structural basis and dipole–dipole interactions will play at most a minimal role. In some situations however, the glass is a disordered array of dipole orientations [10.30]. This sort of state is most likely to occur at very low temperatures in materials that possess dipoles occupying the sites of a regular lattice. At high temperatures the dipole orientations will be disordered but, as the temperature is reduced to low values, each individual

dipole would be expected to adopt their lowest-energy orientation, resulting in a state of ordered dipole orientation. A *dipole glass* will result instead when the dipole–dipole interactions produce forces that generate barriers to the local reorientation and frustrate the ordering process at temperatures low enough that the barriers generated cannot be overcome in any conceivable time.

Ferroelectric Transition

The dielectric response of ferroelectrics at temperatures in the vicinity of their Curie (critical) temperature also exhibit relaxation frequencies that approach zero, just as their dielectric increment approaches infinity (Fig. 10.9a,b) as discussed in Sect. 10.3.1. In this case both the dielectric increment and the relaxation frequency are functions of a hidden variable that characterises the system, the correlation length ξ of the dipole fluctuations. Just as the dielectric increment increases with a power of the correlation length, the relaxation frequency will decrease. Put simply the more dipoles are correlated in the fluctuation the longer the time that is required for its relaxation. Scaling theory [10.33] describes the system by a hierarchy of self-similar correlations. The strongest correlations are between the dipole and its nearest neighbours. This gives a local geometrical arrangement of correlations. The next-strongest correlations are between the same geometrical arrangement of groups of nearest neighbours, and the next strongest is between the same geometry of groups of groups. Eventually the whole system up to the correlation length is constructed in this way. Because the geometrical arrangement is preserved at each stage the properties for each stage have to be proportional to a power of the size. This gives

$$\gamma \propto |(T - T_c)|^\beta \propto \xi^{-\beta/\delta} \quad (10.40)$$

and using (10.31) the relationship

$$\chi_0 \propto (\gamma)^{-\alpha/\beta} \quad (10.41)$$

follows. But the theory can go further and predict the frequency dependence of $\chi_r''(\omega)$ and $\chi_r'(\omega)$ for $\omega > \gamma$. This follows because we can think of the response of the system to a field of frequency ω as being due to the correlation scale that can relax at the frequency ω , i. e. $\xi_\omega \propto \omega^{-\delta/\beta}$. The dielectric increment appropriate to this length scale can be obtained from (10.31) as $(\xi_\omega)^{\alpha/\delta}$ and hence,

$$\chi_r''(\omega) \propto \chi_r'(\omega) \propto (\xi_\omega)^{\alpha/\delta} \propto \omega^{-\alpha/\beta} = \omega^{n-1} \quad (10.42)$$

with $n = 1 - \alpha/\beta$. These relationships describe what is called dynamic scaling. A detailed description of the derivation of the frequency dependence in (10.42) is given in [10.33] and the relationship of the power-law frequency dependence to self-similar relaxations is shown explicitly using scaled electrical circuits in [10.53]. Because of the small range of temperatures around T_c over which the power-law relationships (10.40, 10.41) are expected to hold it is difficult to determine the exponents α and β with any

accuracy, however the dynamic scaling law can be easily demonstrated by using the master-curve technique and determining the locus of the representative point (see the beginning of Sect. 10.3.2). An example is shown in Fig. 10.9c where it can be seen that the representative-point locus of χ_0 as a function of γ is a power law with the same exponent as that of the frequency dependence of $\chi''(\omega)$ and $\chi'_t(\omega)$ for $\omega > \gamma$.

The Information Content of the Relaxation Frequency (Time)

The typical starting point in investigating the relaxation frequency (time) is to make an Arrhenius plot, i. e. $\log(\gamma)$ is plotted as a function of $1/T$. A straight line is taken to indicate an activated process of the form of (10.36), with the gradient yielding the activation energy $\Delta\gamma$. However in most cases data is only available from a restricted temperature range and so the variation of $k_B T/h$ with temperature will be small. In this case the free energy expression (10.38) would give an equally good straight line. Since there are many situations in which the pre-exponential factor in (10.36) will be less than a quantum vibration frequency $10^{12} \text{ Hz} \leq \nu \leq 3 \times 10^{13} \text{ Hz}$, it is difficult to discriminate between the potential-barrier situation and the free-energy barrier in which $S^\ddagger < 0$. Only when $A \gg 3 \times 10^{13} \text{ Hz}$ will identification with relaxation via a free-energy barrier be certain. As can be seen from the foregoing it is difficult to obtain any definite information from the pre-exponential factor in an activated relaxation, except for the latter case where an evaluation of S^\ddagger is possible. Where S^\ddagger can be determined it gives us a qualitative picture of a relaxing dipole centre which has to be structurally distorted in order for the dipole to adopt a different orientation. The corresponding value of H^\ddagger gives the amount of energy that has to be supplied to the group in order for the reorientation to occur even allowing for the distortion. What cannot be determined from this information is the size of the group that is involved, though the behaviour and magnitude of χ_0 may give an idea as to the dipole magnitude and density. To give an idea of the difficulty of a molecular interpretation let us take the case of cyclo-hexanol [10.54]. This molecule forms a plastic crystal, i. e. the molecular centres are located on a crystal lattice but their orientations are disordered. The dipole moment is associated with the only strong polar feature, the alcohol ($-\text{OH}$) group, which can take one of two orientations with respect to the molecule. It would therefore be expected that the measured activation energy of $\approx 0.5 \text{ eV}$ would be the potential barrier to the transfer of the alcohol group between the two positions, with its contributions from the neigh-

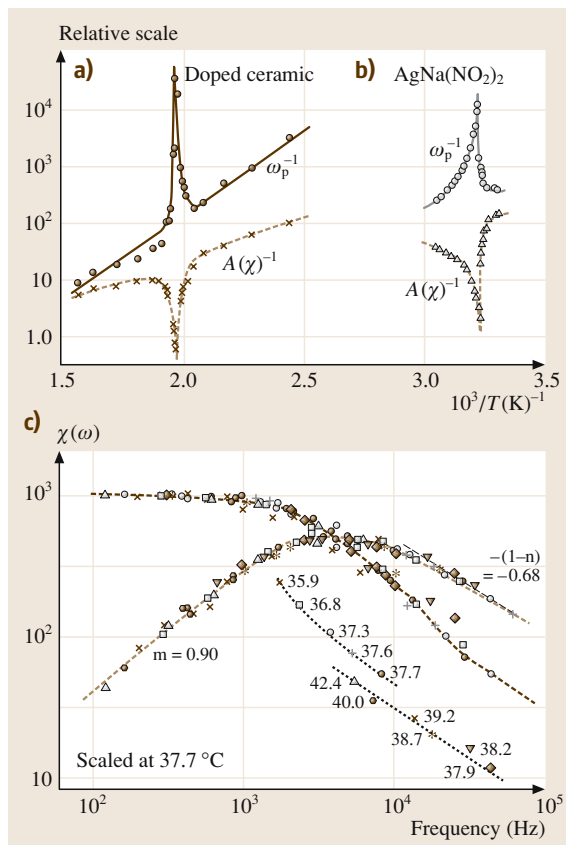


Fig. 10.9a–c An Arrhenius plot of the inverse of the susceptibility increment $A(\chi)(\equiv \chi_0)$ and loss peak frequency $\omega_p(\equiv \gamma)$ is shown for (a) a ceramic and (b) $\text{AgNa}(\text{NO}_2)_2$. The frequency-dependent susceptibility of $\text{AgNa}(\text{NO}_2)_2$ is shown in (c) together with the locus of the representative point as $T_c = 37.8^\circ\text{C}$ is approached from above and below. Note the power-law relationship between the dielectric increment and the relaxation frequency, and the parallelism between this locus and the frequency dependence at $\omega > \gamma$, demonstrating an identical power-law behaviour. After [10.35]

bouring molecular cage as well as the molecular energy change. However, the molecular structure of the cyclohexanol can itself exist in two conformations, the chair and the boat, and can rearrange its orientation in a lattice by passing through the alternative conformation as an intermediate. The free-energy barrier to this interconversion is also ≈ 0.5 eV. So we cannot decide from the relaxation frequency whether the relaxation involves just —OH group transfer or transfer via a boat-to-chair transition or a mixture of both. In this case the pre-exponential frequency $A_{\text{eff}} \approx 6 \times 10^{16}$ Hz, so relaxation has to have a positive S^\ddagger and involve a number of atoms rather than a dipole reorienting on a rigid potential surface.

In the case of near-crystalline materials where the dipole is associated with defect centres we would expect the potential-barrier approach to be the best, but even here the fact that local reorientation is possible implies some sort of interaction between the surroundings and the moving dipole head. Calculations based on a rigid cage should (see for example [10.55]) however be possible, and comparison with experiment can be expected to determine how well this represents the situation and to what extent the transit of the barrier is affected by the barrier friction. Information provided by the dielectric increment should be of help here. The calculation ought to be able to yield an estimate of the reorientable component of the dipole, and if as seems likely the alternative orientations are at different energies, the temperature dependence of χ_0 should follow (10.34).

Although it is conceptually simple to think of dipoles relaxing upon a potential surface that remains unchanged during the relaxation, this is likely to be only an approximation to reality. The fact that alternative orientations exist indicates that in most cases the surrounding structure must be modified to some extent to accommodate the change; at the very least we can expect the dipole to polarise its surroundings differently according to its orientation. The expressions in Sect. 10.3.2 for the relaxation frequency of dipoles relaxing on a free-energy surface and dipoles in ferroelectrics reflect this fact in different ways. The ferroelectric behaviour described in Sect. 10.3.2 shows that when the *dipoles* become extensively correlated the relaxation frequency reduces as an inverse power of the correlation length and the dielectric increment increases as a power of the relaxation length. The self-similar scaling relates this behaviour to the frequency dependence of the susceptibility. The behaviour of the α -response of a glass-forming system involves dipole–structure interaction in a different way. The relaxation frequency approaches zero as T approaches T_g from above. Whatever the details of

the process this behaviour has to indicate an increased difficulty for the dipole to reorient, which here is associated with structural ordering, densification, and atomic packing, rather than long-range correlations as in ferroelectrics. This response is also one for which the dielectric increment is often insensitive to temperature. If we put the two dielectric response features together we come to a picture in which the electric field effectively modifies the configuration entropy of the system in generating a net dipole density fluctuation. The net dipole density produced is essentially the same at different temperatures, so the change in configuration density generated by the electric field does not vary, but the relaxation time increases as the activation enthalpy H^\ddagger and entropy S^\ddagger increases. Put together with the fact that in structural glass formation small local regions are attempting to adopt a crystalline structure, this data indicates that there are *local* values of ground-state configuration entropy that reduce as T_g is approached, with a transition state involving a disordering of the local regions to free the dipole enough to let it adopt an alternative orientation in an equally ordered but different configuration. The dipole density fluctuations that couple to the electric field seem to involve reorganisations of the structure that can occur without a change in the value of the heat content H , i.e. they are essentially configuration entropy fluctuations rather than thermal fluctuations. In contrast to the ferroelectric situation the slowing down of the relaxation is not caused by longer-range correlations but by the increasingly larger numbers of molecular adjustments required to achieve a dipole reorientation.

The message of this section is that in most cases a detailed molecular description of the dipole motions is generally not accessible just from an analysis of the dielectric response. The reason is that, in general, dipole reorientations involve adjustments in the surrounding molecules/atoms that are not easy to define in molecular terms. However by putting together the behaviour of the dielectric increment and relaxation frequency it should be possible to obtain some general idea as to the extent of the connection of the reorientation to the molecular environment and the way in which it takes place.

10.3.3 The Relaxation Peak Shape

The explanation of the frequency dependence of the susceptibility is currently the most contentious of the features of the dielectric response. Many workers are content with just defining the shape by one or other of the empirical functions mentioned in Sect. 10.2.2, or

through the power-law exponents of (10.20, 10.21). This gives a fingerprint of the dipole dynamics but no more. In particular it does not provide a description of the equation of motion of the dipole density fluctuation. Others determine what is termed a *distribution of relaxation times* for the loss peak in $\chi_r''(\omega)$. Essentially this approach is predicated on the assumption that the broadening of the loss peak compared to that of the Debye response (10.19) is the result of dipoles of the same type and dipole moment that each relax according to the Debye equation of motion (10.17) but possess different relaxation times with a distribution denoted by $g(\tau)$, which is defined via (10.43)

$$\chi_r''(\omega) = \int_0^{\infty} g(\tau) \frac{\chi_0 \omega \tau}{1 + \omega^2 \tau^2} d\tau. \quad (10.43)$$

This construction is still no more than a fingerprint unless a physical reason for the distribution $g(\tau)$ can be found. Usually this is ascribed to a distribution of local activation energies associated with dipoles that each exist in their own potential surface independent of one another. The system usually quoted as an example is that of the β -response in the rigid glassy phase, which typically has a very broad loss peak. In this case it is assumed that each dipole that can reorient to contribute to the β -response is essentially trapped in a local potential surface that is held rigid in the glass state. Of course the potential surface is not truly rigid, molecular/atomic vibrations must take place, but it is assumed that their effect on the potential averages out during the relaxation and their only effect is to raise the energy state of the reorienting dipole head to the state at the top of the barrier. One problem associated with this explanation of the origin of $g(\tau)$ is that, if the function $g(\tau)$ is independent of temperature the values of exponents n and m (10.20), (10.21) will be temperature dependent. This does not seem to be the case in general, with these exponents usually either constant or changing at the most slowly or discretely at a transition of state (see for example [10.8, 56]), but there is no real agreement on this point. Of course a temperature-dependent distribution function $g(\tau)$ may be assumed, but then the question arises as to why it is temperature dependent in a system that is presumed to be macroscopically rigid. Another facet of the problem associated with non-Debye loss peaks that does not seem to have received any consideration is the possibility that the magnitude of the reorientable dipole moment associated with each site of a given activation energy is also distributed. It is clear that this is highly likely even if the dipole mo-

ment that changes direction is the same everywhere, as illustrated in Fig. 10.7. Also, as described in Sects. 10.2 and 10.3, the local dipole may be correlated with other dipoles or its surroundings, and in this case we can expect the Debye rate equation not to hold. The ferroelectric result (Sect. 10.3.2) already shows that this is the case when the dipoles motions are correlated giving the system a scale relationship in its dynamics, and even correlation between the dipole and its surroundings, for which there is considerable evidence (Sects. 10.2 and 10.3) can be expected to alter the form of the equation of motion from (10.17), by for example anharmonic coupling between the various modes. Even if we assume that all the criteria for the application of (10.43) are met, the $g(\tau)$ that are required to fit the experimental form of response defined by (10.20), (10.21) [and its corresponding theoretical response function, (10.22)] have unique features that require a physical justification, i. e. there is a cusp or sharp peak at the value of τ corresponding to the characteristic frequency ($\tau_c = 1/\gamma$), and power-law wings to either side whose power exponents are $1 - n[g(\tau) \propto \tau^{1-n}]$ for $\tau < \tau_c$, and $-m[g(\tau) \propto \tau^{-m}]$ for $\tau > \tau_c$. In the Debye case the distribution becomes a delta function at the characteristic relaxation time. Essentially the *distribution of relaxation times* approach is convenient but it is not as easy to justify as would seem at first sight.

The Williams and Watt response function [10.28] started life as a heuristic suggestion but has received some later theoretical support [10.57–62]. The dynamic scaling behaviour appropriate to ferroelectrics gives a clue as to the way in which a frequency-dependent susceptibility of the form of (10.20) can come about, which results from both the equation of motion (10.22) and the response function (10.26). Essentially there has to be a self-similarity (or scaling) between the relaxation frequency of subcomponents of the system and their contribution to the dielectric increment (as illustrated in the circuit model of [10.8]). The theory proposed by Palmer et al. [10.60] refers this scaling to the removal of a hierarchy of constraints, thus for example we may imagine that close neighbours move quickest and remove the constraints imposed on larger groups of molecules and so on. This picture would be appropriate to a system such as a glass-forming material. The assumption however is that the motions are overdamped at all levels of the hierarchy, and hence no bridge is provided to the oscillatory motions known to occur at times close to quantum vibrations. A rather different stochastic approach has been taken by Weron and Jurlewicz [10.61, 62] who assumed that the system re-

relaxation followed a path in which the fastest dipoles out of a distribution relaxed first and then the fastest out of the residual distribution and so on. The key feature is that the relaxing dipole is the extreme fastest from the distribution existing at the time. It was argued that the extreme-value statistical distribution function then led automatically to the response function of (10.26). The choice of appropriate extreme-value distribution was made on the grounds that the relaxation time was a positive definite variable. However this is not a sufficient criterion [10.63]. In order for (10.26) to apply the continuous distribution density of relaxation times (i.e. the distribution the system would have if it were of infinite size) has to be stable to scale changes (see for example [10.64]) and thus has to approach the extreme of long times as an inverse power law, otherwise a different extreme value statistic or *none at all* applies. The required form of distribution from which an extreme selection has to be made is one that applies to the size distribution of scaling systems [10.65] such as percolation clusters [10.66] for example. So even with this stochastic approach we are led back to a system for which the dynamics scale in some way.

The Dissado–Hill function [10.23, 24] for which the response function obeys (10.22) also has scaling features as its basis, however unlike the other approaches it starts with the vibration dynamics of the system. It is assumed that a dipole that can reorient couples local vibration modes to itself. These are no longer extended normal modes but modes centred on the dipole that reduce in frequency according to the molecular mass involved. Their frequencies lie in the region between optical modes and the relaxation frequency and have a scaling relationship one to another. In the theory of Nigmatullin and Le Mehaute [10.67, 68], the modes are impulses that are involved in the dipole relaxation process whose time of action is scaled, i.e. the longer the time of action the more correlated they are to the dipole motion. In general these modes are local versions of coupled optical and acoustic modes and it is not surprising that they extend to such low frequencies as those involved in relaxation, as acoustic modes essentially extend to zero frequency. Their coupling with the dipole leads to the high-frequency power law of (10.20), where n expresses the extent to which the dipole reorientation couples to the surroundings, i.e. $n = 0$ corresponds to no coupling and the dipole moves independently of its surroundings, and $n = 1$ corresponds to full coupling in which the dipole motion is just part of the local mode. In a sense the short-time development of the response function of (10.22) is that of the changes in the con-

figuration entropy as various amounts of different local modes are progressively coupled into the dipole motion [10.23]. In this case there is no necessity for n to be temperature dependent. At the characteristic relaxation frequency, the *characteristic* dipole group relaxes and transfers energy to the heat bath. The low-frequency behaviour of (10.21) is the result of a distribution in the ensemble of locally coupled dipole motions. This occurs because the motions of local dipole centres may be weakly coupled to one another. As a result the relaxation of the centres proceeds in a scaled or self-similar manner. First the dipole in a local centre relaxes with respect to its own environment, this leaves each dipole centre unrelaxed with respect to one another. Next groups of dipole centres, with some arrangement depending upon the specific structure involved, relax as a group. Then groups of groups relax and so on. Each level of inter-group complexity essentially has a time scale associated with its relaxation that cannot be reached until the preceding level has been completed. This is rather similar to the constraint relaxation concept of Palmer et al. [10.60]. The power-law exponent m expresses the way that this hierarchy of relaxing groups is scaled, by defining the power-law tail of the distribution of inter-group relaxation times in the ensemble [10.69]. A value of $m = 1$ corresponds to a sequence of inter-group relaxations with a relaxation time that is proportional to the number of groups involved in the sequence [10.69]. This implies that the sequential events are uncorrelated, i.e. the long-time relaxation is a white-noise (random) process [10.23]. When combined together with $n = 0$ the Debye response is recovered. On the other hand a value of m approaching zero corresponds to relaxation times that are a very high power of the number of groups involved [10.69] and indicates a very strong connection between groups at all levels of the hierarchy. This will spread the response to very low frequencies, as observed. Essentially m is a measure of the extent to which energy is transferred to the heat bath (dissipated) at each level of the hierarchy compared to being stored in the inter-group interactions of the next level. Again scaling is at the basis of the theory, but now with two different ways in which it can be involved. This theory is not generally accepted. The controversial parts of the theory are firstly the coupling of the dipole motions with vibration modes, which modifies the oscillator behaviour towards an over-damped form, and secondly the hierarchy of relaxations whereby energy is transferred to the heat bath. However it should be noted that the susceptibility function that results has a general form that agrees well with experiment. In addition the concepts are reasonable given

the complexity that is likely to occur in the internal motions once an ideal crystalline regularity is ruled out by the possibility of dipole reorientation. Thus for example this concept would apply to the dipoles involved in the β -response of the glass state as well as correlated motions of dipoles over long distances, since even in a macroscopically rigid material local vibrations take place. In fact the limited regions of local order in a glass phase can be expected to favour such local modes and increase the coupling of the dipole motions with them, as observed.

The Information Content of the Loss Peak Shape

It is clear from the foregoing discussion that for all theoretical models of the loss peak shape the characteristic or loss peak frequency is but the culmination of a process in which subsections of the dipole and environment (with or without dipoles) are mixed into the motion of the dipole centre. In these models the dipole is not an independent entity, but rather an entity that is connected to some extent over a region that may be small or large. This implies that the dipole is not a particle that relaxes on a rigid potential surface independently of its environment. Only the *distribution of relaxation times* approach preserves the latter concept. If the theoretical models are correct they reflect the fact that we are looking at entities that are not truly of molecular scale but are of a mesoscopic nature. The correlations noted to occur in χ_0 and the need to use free energy rather than potential surfaces in describing the relaxation frequency

support this view. The local entities involved are however not rigid features like permanent dipoles, and for this reason we should expect there to be weak connections between them that can be expected to relate to the way in which the relaxation of the whole system takes place. That is, not all entities relax at the characteristic time. As one entity relaxes its neighbours have to come into equilibrium with its new orientation and the system approaches equilibrium more slowly [i.e. as the time power law $t^{-(1+m)}$] than the exponential behaviour of the Debye response function or the expanded exponential function. The information contained in the loss peak shape indicates the way in which a dipole density fluctuation evolves from its state when initially created to an ensemble of mesoscopic dipole centres. The broadening of the peak from that of a Debye peak indicates the involvement of faster and slower processes as part of the overall mechanism, whatever their detailed origin, and in particular processes that have a scale relationship to one another. This must apply even to a distribution of relaxation times because of the unique form required for that distribution. Equation (10.22) implies an equivalent description that refers the overall relaxation process to a conversion of the vibration oscillation at short times to an overdamped motion as the dipole density fluctuation dissipates its energy irreversibly. In this sense evaluation of the shape parameters n and m give a means of describing this conversion process. At the very least they give a sense of the scaling involved in spreading the relaxation process around the characteristic relaxation frequency or equivalently the characteristic relaxation time.

10.4 Charge Transport

All dielectrics possess a constant (DC) conductivity (σ_{DC}), although usually it is very weak. Since $\chi''(\omega) = \sigma(\omega)/\omega$ as demonstrated in Sect. 10.1 (10.10), it would be expected that a dielectric response at low frequencies ($f \lesssim 10^{-2}$ Hz) would take a form in which $\chi''(\omega) = \sigma_{DC}/\omega$ and χ' is independent of frequency. In many cases however the conduction process is blocked at the electrodes or internal interfaces. In this case the DC conduction charges the interface, which behaves as a capacitor, and the whole system behaves as a single dipole. As long as the interface does not possess relaxation dynamics of its own, the response that would be observed is that given by the Debye response of (10.18, 10.19), with $\tau = 1/\gamma = RC_i$, where R is the resistance of the body of the material and C_i is the

capacitance of the interface. The measured dielectric increment $\chi_0 = dC_i/A$, where A is the electrode area and d is the sample thickness, and can be very large depending upon the ratio of the sample thickness to that of the interface. The situation where the interface has a frequency-dependent capacitance has been thoroughly discussed by Jonscher [10.6] who has shown that $\chi_r'(\omega)$ is modified from $\chi_r'(\omega) \propto \omega^{-2}(\omega > \gamma)$ to $\chi_r'(\omega) \propto \omega^{-q}$, while $\chi_r''(\omega) \propto (1/\omega)(\omega > \gamma)$ as in (10.19). The value of q lies in the range $1 < q < 2$ with its value depending upon the frequency dependence of the interface capacitance.

The bulk DC conductance arises from charged particles whose movements are not bound to a charge of the opposite polarity as in dipoles but are free to

move independently of their countercharge, resulting in a net charge displacement in the same way that a liquid flows. However the transport of charged particles within the body of the sample can give rise to a very different form of response when their movement lies along defined paths such that the longer the displacement of the charge the lower the number of paths or equivalently the more difficult the transport becomes. This behaviour was called low-frequency dispersion by *Jonscher* [10.6, 20] and quasi-DC conduction (q-DC) in the theoretical model of *Dissado and Hill* [10.70] who wished to distinguish it from low-frequency dipole responses. At frequencies below some characteristic value ω_c this form of response takes the form,

$$\sigma(\omega) \propto \omega^{1-p}$$

$$\left[\text{i. e. } \chi_r''(\omega) \propto \chi_r'(\omega) \propto \omega^{-p} \right], \quad \omega < \omega_c \quad (10.44)$$

and at frequencies above ω_c ,

$$\sigma(\omega) \propto \omega^n$$

$$\left[\text{i. e. } \chi_r''(\omega) \propto \chi_r'(\omega) \propto \omega^{n-1} \right], \quad \omega > \omega_c. \quad (10.45)$$

The power-law exponents p in (10.44) and n in (10.45) have positive fractional values near to unity. It is obviously difficult to identify a value of p close to unity from measurements of $\chi_r''(\omega)$ [or equivalently $\sigma(\omega)$] and in many cases it is assumed that the measured behaviour shows a static (DC) conductivity. It is then common to subtract its supposed value from the measured data for $\sigma(\omega)$ to obtain an expression for the dipole relaxation response supposedly responsible for the behaviour at $\omega > \omega_c$. The values obtained in this way for $\chi_r''(\omega)$ at frequencies $\omega < \omega_c$ will not be zero as $\sigma(\omega)$ is not in fact constant, instead they will reduce as the frequency is reduced. This procedure yields a spurious loss peak in $\chi_r''(\omega)$ if the response is actually due to the q-DC mechanism, for which the high-frequency behaviour is an essential component of the whole q-DC mechanism and can never be resolved as a separate peak in $\chi_r''(\omega)$. The way to be certain that the response is really of the q-DC form is to measure the frequency response for $\chi_r'(\omega)$ and show that it takes the same frequency dependence. A convenient check is to determine the ratio of $\chi_r''(\omega)$ to $\chi_r'(\omega)$ (i. e. $\tan \delta$) which will have a constant value [10.6, 20] given by

$$\chi_r''(\omega)/\chi_r'(\omega) = \tan \delta = \cot[(1-p)\pi/2]. \quad (10.46)$$

Here $\tan \delta$ is called the loss tangent and δ is the phase angle between the real and imaginary components of the

susceptibility. This relationship holds for pairs of values of $\chi_r'(\omega)$ and $\chi_r''(\omega)$ at the same frequency even if the measurements are noisy and so make it difficult to determine accurately the value of p from the frequency dependence. Another situation where it is difficult to detect the q-DC behaviour occurs in heterogeneous materials when one component has a low DC conductivity. This will add to the AC component, (10.44) and obscure the q-DC behaviour. In this case the DC conductivity can be eliminated from the data, if it is available over a large enough frequency range, by applying the Kramers–Kronig transform of (10.25) to obtain the function $\chi_r''(\omega)$ without the DC component (σ_{DC}/ω). The validity of the procedure can be checked by applying the inverse transform (10.24) to the measured data for $\chi_r''(\omega)$. This should yield the measured $\chi_r'(\omega)$ since the DC conductivity does not contribute to the real component of the susceptibility.

The q-DC behaviour, (Fig. 10.10a), is most often found in materials that are heterogeneous on a mesoscopic scale such as ceramics [10.71], rocks [10.72], porous structures [10.73], and biological systems [10.74]. In these materials charged particles are transported via structured paths over some finite range. The transported charge and its countercharge give rise to an effective dipole with a large dipole moment. However the q-DC behaviour rarely appears as an isolated response. Because of the heterogeneous nature of the materials it is usually found to be electrically in series with other dielectric response elements such as interface capacitances, and electrically in parallel with a capacitive circuit element. The origin of the q-DC behaviour lies in a hidden scale relationship, with the dipole contribution to the susceptibility increment and its relaxation time both being a power of the length over which the transport takes place. The circuit models of *Dissado and Hill et al.* [10.8, 75] show how this behaviour can be produced when the system is represented by a geometrically self-similar arrangement of transport paths and blocking capacitive regions. Such geometrical regularity is not essential however [10.76]; a random arrangement of conductors (transport paths) in a dielectric (i. e. residual set of capacitances) will also result in the q-DC behaviour. It is clear that this construction yields percolation clusters below the size necessary to span the material, and these sub-percolation clusters will of necessity possess scaling relationships dependent on their size and the number of paths within them, and between clusters of different sizes. Such percolation systems also show q-DC behaviour when below their critical limit [10.77]. The theory proposed in [10.70]

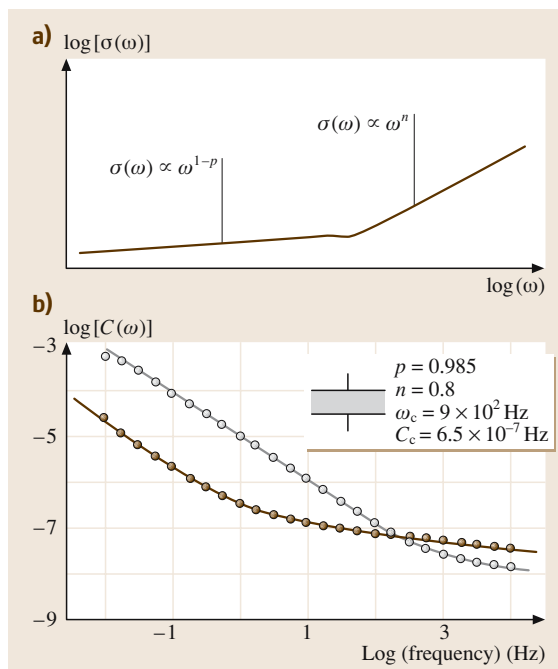


Fig. 10.10 (a) A schematic representation of the frequency dependence of $\sigma(\omega)$ for the q-DC process, (b) the equivalent q-DC susceptibility function $\chi(\omega)/\chi_c \propto (1 + i\omega\tau)^{n-1} {}_2F_1[1-n; 1+p; 2-n; (1+i\omega\tau)^{-1}]$ derived from the response function of (10.47) fitted to data from the leaf of *Nicotiana* (Solanaceae), the inset gives the fitting parameters, with $\omega_c \equiv 1/\tau$, and $C_c \equiv \chi_c$. This data was chosen as it exhibits an isolated q-DC process. As before ${}_2F_1(\cdot; \cdot)$ is the Gaussian hypergeometric function; see [10.25]. After [10.79]

described the q-DC mechanism in similar terms to that used in percolation theory. At high frequencies they assumed that the charged particles had a short range of motion restricted to local clusters in which they were bound to their countercharges. The power-law exponent n reflected the extent of binding of the motions of the charges in the cluster and the local motions in the same way as for the dipole motions. Its value close to one indicates that the binding is strong and that the polarisation of the local clusters increases only to a small extent as the frequency is lowered, i.e. the net displacement of positive and negative charge is local and small. At the characteristic frequency the charged particles become free of their locality and are able to move

over much longer distances. The range allowed by the AC frequency increases as the frequency reduces, but the number of effective routes decreases as the distance becomes longer and hence the current becomes smaller. This leads to the weak dependence of the conductivity upon the AC frequency defined in (10.44). In the zero-frequency limit $\sigma(\omega) = 0$, which means that there are no paths across a system of infinite length, i.e. the system is formally below the percolation limit. Of course in reality the sample is of finite size and some routes may cross the sample, leading to a termination of the q-DC behaviour at low frequencies, either by a blocking capacitance at the electrodes or a weak DC current [10.78]. The theory leads to an analytical expression for the q-DC susceptibility in terms of a confluent hypergeometric function [10.25] (see [10.70] for the detailed expression), which is obtained from a response function that obeys an equation of motion similar to that of the permanent dipoles

$$\frac{d^2\phi}{dt^2} + \frac{(2+n+t/\tau)}{t} \frac{d\phi}{dt} + \frac{[n+(t/\tau)(1-p)]}{t^2} \phi = 0. \quad (10.47)$$

At times $t < \tau = 1/\omega_c$ the behaviour of the dipole density fluctuation is the same as that appropriate for permanent dipoles (10.22) with the same limiting solution, i.e. $\phi(t) \propto t^{-n}$ at $t < \tau$. At long times $t > \tau = 1/\omega_c$, however, it takes a different form in which the dipolar fluctuation relaxation has only a weak dependence on t , i.e. $\phi(t) \propto t^{p-1}$ at $t > \tau$. An example of the q-DC behaviour in frequency is given in Fig. 10.10b. The reader is referred to [10.74, 79, 80] for other examples.

In the limit of $p = 1$ [$\sigma(\omega) = \text{constant}$; $\omega < \omega_c$], the last term of (10.47) is zero in the limit $t > \tau$ at which time the decay of ϕ is governed by the first two terms. The solution for $\phi(t)$ now takes the form

$$\begin{aligned} \phi(t) &\propto t^{-n}, & t < \tau, \\ \phi(t) &\propto \exp(-t/\tau), & t > \tau. \end{aligned} \quad (10.48)$$

In this case we can see that the dipole density fluctuations are produced within the local clusters and that they relax by freeing the charges to move independently at $t = \tau$. At this time there is a density of completely free charged particles with charge and countercharge cancelling throughout the system giving a net dipole of zero as in a true DC-conduction process.

10.5 A Few Final Comments

The basic difficulty associated with the interpretation of dielectric responses is that they are of necessity macroscopic measurements made on samples that contain enormous numbers of atoms and molecules. In condensed-phase materials it is not possible to consider these systems as made up of local entities each moving independently of one another. All entities that contribute a permanent dipole are part of the condensed-phase structure, and even though they have a degree of freedom associated with the possibility of dipole reorientation, they will have motions that are correlated or connected to some extent to the molecules/atoms in their environment. This means that dipole reorientation is not that of a bare entity; instead it involves to some extent a local region. These regions will behave differently in different kinds of material and their definition and the way of describing their behaviour has not yet been established with any sort of rigour. Since we are dealing

with a macroscopic measurement, there will of necessity be an ensemble of the local entities. This will result in a distribution of entities, but since these are part of the structure there will be some sort of connection between them unless the structure itself is disconnected dynamically. This means that fluctuations will take place among our entities, and perhaps even dissociation and amalgamation. These effects will also have an influence on the relaxation dynamics. In the foregoing I have tried to give some simple pictures as to what is happening and to do so in a holistic way by correlating information from different facets of the measurement. What is abundantly clear is that the dynamics of such systems are very complicated in detail, but I hope that I have done enough to convince you that there are some basic features of the relaxation process that are common to all systems of this type, even though a full understanding of their nature does not yet exist.

References

- 10.1 P. Debye: *Polar Molecules* (Dover, New York 1945)
- 10.2 C. P. Smyth: *Dielectric Behaviour and Structure* (McGraw-Hill, New York 1955)
- 10.3 N. G. McCrum, B. E. Read, G. Williams: *Anelastic and Dielectric Effects in Polymeric Solids* (Wiley, New York 1967)
- 10.4 V. V. Daniels: *Dielectric Relaxation* (Academic, New York 1967)
- 10.5 C. J. F. Bottcher, P. Bordewijk: *Theory of Electric Polarisation*, Vol. I, II (Elsevier, Amsterdam 1978)
- 10.6 A. K. Jonscher: *Dielectric Relaxation in Solids* (Chelsea Dielectric, London 1983)
- 10.7 K. L. Ngai, G. B. Wright (Eds.): *Relaxations in Complex Systems*. In: *Proc. The International Discussion Meeting on Relaxations in Complex Systems* (Elsevier, Amsterdam 1991)
- 10.8 R. M. Hill: Electronic Materials from Silicon to Organics. In: *Dielectric properties and materials*, ed. by L. S. Miller, J. B. Mullin (Plenum, New York 1991) pp. 253–285
- 10.9 T. Furukawa: IEEE Trans. E.I. **24**, 375 (1989)
- 10.10 C. J. F. Bottcher: *Theory of Electric Polarisation* (Elsevier, Amsterdam 1952) p. 206
- 10.11 M. Born, E. Wolf: *Principles of Optics* (Pergamon, Oxford 1965)
- 10.12 A. R. von Hippel (ed.): *Dielectric Materials and Applications* (Wiley, New York 1958)
- 10.13 R. N. Clarke, A. Gregory, D. Connell, M. Patrick, I. Youngs, G. Hill: Guide to the Characterisation of Dielectric Materials at RF and Microwave Frequencies. In: *NPL Good Practice Guide* (Pub. Inst. Measurement and Control, London 2003)
- 10.14 W. Heitler: *The Quantum Theory of Radiation*, 3rd edn. (Dover, London 1984)
- 10.15 D. Bohm, D. Pines: Phys. Rev. **82**, 625 (1951)
- 10.16 H. Eyring, J. Walter, G. E. Kimball: *Quantum Chemistry* (Wiley, New York 1960)
- 10.17 P. Wheatley: *The Determination of Molecular Structure* (Clarendon, Oxford 1959) Chap. XI
- 10.18 R. Kubo: Rep. Prog. Phys. **29**, 255–284 (1966)
- 10.19 R. Kubo: J. Phys. Soc. Jpn. **12**, 570 (1957)
- 10.20 A. K. Jonscher: J. Phys. D Appl. Phys. **32**, R57 (1999)
- 10.21 S. Havriliak, S. Negami: J. Polym. Sci. **C14**, 99 (1966)
- 10.22 S. Jr. Havriliak, S. J. Havriliak: *Dielectric and Mechanical Relaxation in Materials* (Hanser, New York 1997)
- 10.23 L. A. Dissado, R. M. Hill: Proc. R. Soc. London **390**(A), 131 (1983)
- 10.24 L. A. Dissado, R. R. Nigmatullin, R. M. Hill: Dynamical Processes in Condensed Matter. In: *Adv. Chem. Phys.*, Vol. LXIII, ed. by R. Evans M. (Wiley, New York 1985) p. 253
- 10.25 M. Abramowitz, I. A. Stegun: *Handbook of Mathematical Functions* (Dover, New York 1965)
- 10.26 L. A. Dissado, R. M. Hill: Chem. Phys. **111**, 193 (1987)
- 10.27 L. A. Dissado, R. M. Hill: Nature (London) **279**, 685 (1979)
- 10.28 G. Williams, D. C. Watt: Trans. Farad. Soc. **66**, 80 (1970)
- 10.29 R. Kohlrausch: Pogg. Ann. Phys. **91**, 198 (1854)

- 10.30 A. K. Loidl, J. Knorr, R. Hessinger, I. Fehst, U. T. Hochli: *J. Non-Cryst. Solids* **269**, 131 (1991)
- 10.31 C. Kittel: *Introduction to Solid State Physics* (Wiley, New York 1966)
- 10.32 J. Joffrin, A. Levelut: *J. Phys. (Paris)* **36**, 811 (1975)
- 10.33 P. C. Hohenberg, B. I. Halperin: *Rev. Mod. Phys.* **49**, 435–479 (1977)
- 10.34 J. G. Kirkwood: *J. Chem. Phys.* **7**, 911 (1939)
- 10.35 L. A. Dissado, R. M. Hill: *Phil. Mag. B* **41**, 625–642 (1980)
- 10.36 L. A. Dissado, M. E. Brown, R. M. Hill: *J. Phys. C* **16**, 4041–4055 (1983)
- 10.37 L. A. Dissado, R. M. Hill: *J. Phys. C* **16**, 4023–4039 (1983)
- 10.38 H. A. Kramers: *Physica* **VII**(4), 284–304 (1940)
- 10.39 R. F. Grote, J. T. Hynes: *J. Chem. Phys.* **73**, 2715–2732 (1980)
- 10.40 W. A. Phillips: *Proc. R. Soc. London A* **319**, 535 (1970)
- 10.41 J. le G. Gilchrist: *Cryogenics* **19**, 281 (1979)
- 10.42 J. le G. Gilchrist: Private communication with R. M. Hill, reported in 10.44, (1978)
- 10.43 H. Eyring: *J. Chem. Phys.* **4**, 283 (1936)
- 10.44 R. M. Hill, L. A. Dissado: *J. Phys. C* **15**, 5171 (1982)
- 10.45 S. H. Glarum: *J. Chem. Phys.* **33**, 1371 (1960)
- 10.46 C. A. Angell: *Encyclopedia of Materials*. In: *Science and Technology*, Vol. 4, ed. by K. H. J. Buschow, R. W. Cahn, M. C. Fleming, B. Ilshner, E. J. Kramer, S. Mahajan (Elsevier, New York 2001) pp. 3565–3575
- 10.47 W. Kauzmann: *Chem. Rev.* **43**, 219 (1948)
- 10.48 G. P. Johari, M. Goldstein: *J. Chem. Phys.* **53**, 2372 (1970)
- 10.49 M. L. Williams, R. F. Landel, J. D. Ferry: *J. Am. Chem. Soc.* **77**, 3701 (1955)
- 10.50 M. Goldstein: *J. Chem. Phys.* **39**, 3369 (1963)
- 10.51 D. Turnbull, M. H. Cohen: *J. Chem. Phys.* **14**, 120 (1961)
- 10.52 R. R. Nigmatullin, S. I. Osokin, G. Smith: *J. Phys. Cond. Matter* **15**, 1 (2003)
- 10.53 R. M. Hill, L. A. Dissado, R. R. Nigmatullin: *J. Phys. Cond. Matter* **3**, 9773 (1991)
- 10.54 M. Shablakh, L. A. Dissado, R. M. Hill: *J. Chem. Soc. Faraday Trans. 2* **79**, 369 (1983)
- 10.55 R. Pirc, B. Zeks, P. Goshar: *Phys. Chem. Solids* **27**, 1219 (1966)
- 10.56 K. Pathmanathan, L. A. Dissado, R. M. Hill: *Mol. Cryst. Liq. Cryst.* **135**, 65 (1986)
- 10.57 K. L. Ngai, A. K. Jonscher, C. T. White: *Nature* **277**, 185 (1979)
- 10.58 K. L. Ngai, A. K. Rajgopal, S. Tietler: *J. Phys. C* **17**, 6611 (1984)
- 10.59 K. L. Ngai, R. W. Rendell, A. K. Rajgopal, S. Tietler: *Ann. Acad. Sci. NY* **484**, 150 (1986)
- 10.60 R. G. Palmer, D. Stein, E. S. Abrahams, P. W. Anderson: *Phys. Rev. Lett.* **53**, 958 (1984)
- 10.61 K. Weron: *J. Phys. Cond. Matter* **4**, 10507 (1992)
- 10.62 K. Weron, A. Jurlewicz: *J. Phys. A* **26**, 395 (1993)
- 10.63 E. J. Gumbel: *Statistics of Extremes* (Columbia University Press, New York 1958)
- 10.64 J. T. Bendler: *J. Stat. Phys.* **36**, 625 (1984)
- 10.65 J. Klafter, M. F. Schlesinger: *Proc. Natl. Acad. Sci.* **83**, 848 (1986)
- 10.66 D. Stauffer: *Introduction to Percolation Theory* (Taylor Francis, London 1985)
- 10.67 R. R. Nigmatullin: *Theor. Math. Phys.* **90**, 354 (1992)
- 10.68 R. R. Nigmatullin, A. Le Mehaute: *Int. J. Sci. Geores.* **8**, 2 (2003)
- 10.69 L. A. Dissado, R. M. Hill: *J. Appl. Phys.* **66**, 2511 (1989)
- 10.70 L. A. Dissado, R. M. Hill: *J. Chem. Soc. Faraday Trans. 1* **80**, 325 (1984)
- 10.71 T. Ramdeen, L. A. Dissado, R. M. Hill: *J. Chem. Soc. Faraday Trans. 2* **80**, 325 (1984)
- 10.72 R. R. Nigmatullin, L. A. Dissado, N. N. Soutougin: *J. Phys. D* **25**, 113 (1992)
- 10.73 A. Puzenko, N. Kozlovich, A. Gutina, Yu. Feldman: *Phys. Rev. B* **60**, 14348 (1999)
- 10.74 L. A. Dissado: *Phys. Med. Biol.* **35**, 1487 (1990)
- 10.75 L. A. Dissado, R. M. Hill: *Phys. Rev. B* **37**, 3434 (1988)
- 10.76 D. P. Almond, C. R. Brown: *Phys. Rev. Lett.* **92**, 157601 (2004)
- 10.77 Yu. Feldman, N. Kozlovich, Yu. Alexandrov, R. Nigmatullin, Ya. Ryabov: *Phys. Rev. E* **54**, 20–28 (1996)
- 10.78 L. A. Dissado, R. M. Hill, C. Pickup, S. H. Zaidi: *Appl. Phys. Commun.* **5**, 13 (1985)
- 10.79 R. M. Hill, L. A. Dissado, K. Pathmanathan: *J. Biol. Phys.* **15**, 2 (1987)
- 10.80 M. Shablakh, L. A. Dissado, R. M. Hill: *J. Biol. Phys.* **12**, 63 (1984)

11. Ionic Conduction and Applications

Solid state ionic conductors are crucial to a number of major technological developments, notably in the domains of energy storage and conversion and in environmental monitoring (such as battery, fuel cell and sensor technologies). Solid state ionic membranes based on *fast ion conductors* potentially provide important advantages over liquid electrolytes, including the elimination of sealing problems and the ability to miniaturize electrochemical devices using thin films. This chapter reviews methods of optimizing ionic conduction in solids and controlling the ratio of ionic to electronic conductivity in mixed conductors. Materials are distinguished based on whether they are characterized by intrinsic vs. extrinsic disorder, amorphous vs. crystalline structure, bulk vs. interfacial control, cation vs. anion conduction and ionic vs. mixed ionic–electronic conduction. Data for representative conductors are tabulated.

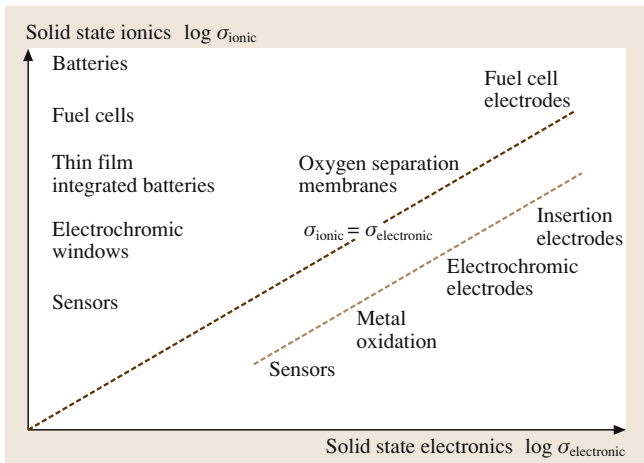
A number of applications that rely on solid state electrolytes and/or mixed ionic–electronic conductors are considered, and the criteria used to choose such materials are reviewed. Emphasis is placed on fuel cells, sensors and batteries,

The ionic bonding of many refractory compounds allows for ionic diffusion and correspondingly, under the influence of an electric field, ionic conduction. This contribution, for many years, was ignored as being inconsequential. However, over the past three to four decades, an increasing number of solids that support anomalously high levels of ionic conductivity have been identified. Indeed, some solids exhibit levels of ionic conductivity comparable to those of liquids. Such materials are termed *fast ion conductors*. Like solid state electronics, progress in *solid state ionics* has been driven by major technological developments, notably in the domains of energy storage and conversion and environmental monitoring, based on ongoing developments in battery, fuel cell and sensor technologies. Some of the most important applications of solid state electron-

11.1	Conduction in Ionic Solids	214
11.2	Fast Ion Conduction	216
11.2.1	Structurally Disordered Crystalline Solids	216
11.2.2	Amorphous Solids	219
11.2.3	Heavily Doped Defective Solids....	219
11.2.4	Interfacial Ionic Conduction and Nanostructural Effects	220
11.3	Mixed Ionic–Electronic Conduction	221
11.3.1	Defect Equilibria	221
11.3.2	Electrolytic Domain Boundaries...	222
11.4	Applications	223
11.4.1	Sensors	223
11.4.2	Solid Oxide Fuel Cells (SOFC)	224
11.4.3	Membranes	225
11.4.4	Batteries	225
11.4.5	Electrochromic Windows	226
11.5	Future Trends	226
	References	226

where there is strong scientific and technological interest. The chapter concludes by considering how solid state ionic materials are likely to be used in the future, particularly in light of the trend for miniaturizing sensors and power sources.

ics and solid state ionics, and their categorization by type and magnitude of conductivity (such as dielectric, semiconducting, metallic and superconducting), are illustrated in Fig. 11.1 [11.1]. This figure also emphasizes that solids need not be strictly ionic or electronic, but may and often do exhibit mixed ionic–electronic conductivity. These *mixed conductors* play a critical role – particularly as electrodes – in solid state ionics, and are receiving comparable if not more attention than solid electrolytes at the present. Such solids are the result of a combination of the fields of *solid state ionics* and *solid state electrochemistry*, and they have grown in importance as our society has become more acutely concerned with efficient and environmentally clean methods for energy conversion, conservation and storage [11.2].



Solid state ionic membranes provide important potential advantages over liquids. The most important of these include: (a) elimination of sealing problems

11.1 Conduction in Ionic Solids

The electrical conductivity, σ , the proportionality constant between the current density j and the electric field E , is given by

$$j/E = \sigma = \sum_i c_i Z_i q \mu_i \quad (11.1)$$

where c_i is the carrier density (number/cm³), μ_i the mobility (cm²/Vs), and $Z_i q$ the charge ($q = 1.6 \times 10^{-19}$ C) of the i th charge carrier. The huge (many orders of magnitude) differences in σ between metals, semiconductors and insulators generally result from differences in c rather than μ . On the other hand, the higher conductivities of electronic versus ionic conductors are generally due to the much higher mobilities of electronic versus ionic species [11.3].

Optimized ionic conduction is a well-known characteristic of molten salts and aqueous electrolytes wherein all ions move with little hindrance within their surroundings. This leads to ionic conductivities as high as 10^{-1} – 10^1 S/cm in molten salts at temperatures of 400–900 °C [11.4]. Typical ionic solids, in contrast, possess limited numbers of mobile ions, hindered in their motion by virtue of being trapped in relatively stable potential wells. Ionic conduction in such solids easily falls below 10^{-10} S/cm for temperatures between room temperature and 200 °C. In the following sections, we examine the circumstances under which the magnitude

Fig. 11.1 Illustration of typical applications of ionic and electronic conductors as a function of the magnitude of electrical conductivity. Applications requiring mixed ionic electronic conductivity fall within the quadrant bounded by the two axes. After [11.1]

associated with chemically reactive liquid or molten electrolytes; (b) minimization of discharge under open circuit conditions, and; (c) the ability to miniaturize electrochemical devices through the use of thin films. In the following, we begin by discussing methods of optimizing ionic conduction in solids and controlling the ratio of ionic to electronic conductivity. We then consider a number of applications that rely on solid state electrolytes and/or mixed ionic–electronic conductors and the criteria that should be used when selecting materials. We conclude by considering how solid state ionic materials are likely to be used in the future, particularly in light of trends related to the miniaturization of sensors and power sources.

of ionic conduction in solids approaches that found in liquid electrolytes.

The motion of ions is described by an activated jump process, for which the diffusion coefficient is given by [11.5]

$$\begin{aligned} D &= D_0 \exp(-\Delta G/kT) \\ &= \gamma(1-c)Za^2v_0 \exp(\Delta S/k) \exp(-E_m/kT), \end{aligned} \quad (11.2)$$

where a is the jump distance, v_0 the attempt frequency, and E_m the migration energy. The factor $(1-c)Z$ defines the number of neighboring unoccupied sites, while γ includes geometric and correlation factors. Note that the fractional occupation c here should not be confused with c_i , the charge carrier concentration, nor should the number of nearest neighbors Z be confused with Z_i , the number of charges per carrier defined in (11.1). Since the ion mobility is defined by $\mu_i = Z_i q D_i / k_B T$, where D_i and k_B are the diffusivity and Boltzmann constant respectively, and the density of carriers of charge $Z_i q$ is Nc , where N is the density of ion sites in the sublattice of interest, the ionic conductivity becomes

$$\begin{aligned} \sigma_{\text{ion}} &= \gamma \left[N(Z_i q)^2 / k_B T \right] c(1-c)Za^2v_0 \\ &\quad \times \exp(\Delta S/k_B) \exp(-E_m/k_B T) \\ &= (\sigma_0/T) \exp(-E/k_B T). \end{aligned} \quad (11.3)$$

or

$$\sigma_{\text{ion}} = \gamma N(Z_i q)^2 c(1-c) Z a^2 \gamma_0 / k_B T \times \exp(\Delta S / k_B) \exp(-E_m / k_B T) \quad (11.4)$$

This expression shows that σ_{ion} is nonzero only when the product $c(1-c)$ is nonzero. Since all normal sites are fully occupied ($c = 1$) and all interstitial sites are empty ($c = 0$) in a perfect classical crystal, this is expected to lead to highly insulating characteristics. The classical theory of ionic conduction in solids is thus described in terms of the creation and motion of atomic defects, notably vacancies and interstitials.

Three mechanisms for ionic defect formation in oxides should be considered. These are (1) thermally induced intrinsic ionic disorder (such as Schottky and Frenkel defect pairs), (2) redox-induced defects, and (3) impurity-induced defects. The first two categories of defects are predicted from statistical thermodynamics [11.6], and the latter form to satisfy electroneutrality. Examples of typical defect reactions in the three categories, representative of an ionically bonded binary metal oxide, are given in Table 11.1, in which the $K_i(T)$ s represent the respective equilibrium constant and $a_{\text{N}_2\text{O}_3}$ the activity of the dopant oxide N_2O_3 added to the host oxide MO_2 . Schottky and Frenkel disorder (1, 2) leave the stoichiometric balance intact. Reduction–oxidation behavior, as represented by (3), results in an imbalance in the ideal cation-to-anion ratio and thus leads to *nonstoichiometry*. Note that equilibration with the gas phase, by the exchange of oxygen between the crystal lattice and the gas phase, generally results in the simultaneous generation of both ionic and electronic carriers. For completeness, the equilibrium between electrons and holes is given in (4).

Altrivalent impurities [for example N^{3+} substituted for the host cation M^{4+} – see (5)] also contribute to the generation of ionic carriers, commonly more than intrinsic levels do. This follows from the considerably reduced ionization energies required to dissociate impurity-defect pairs as compared to intrinsic defect generation. For example, E_A might correspond to the energy required to dissociate an acceptor–anion vacancy

pair or E_D to the energy needed to dissociate a donor–anion interstitial pair. Such dissociative effects have been extensively reported in both halide and oxide literature [11.7]. A more detailed discussion is provided below in the context of achieving high oxygen ion conductivity in solid oxide electrolytes.

The oxygen ion conductivity σ_i is given by the sum of the oxygen vacancy and interstitial partial conductivities. In all oxygen ion electrolytes of interest, the interstitial does not appear to make significant contributions to the ionic conductivity, and so it is the product of the oxygen vacancy concentration $[\text{V}_\text{O}^{\bullet\bullet}]$, the charge $2q$, and the mobility (μ_v):

$$\sigma_i \approx [\text{V}_\text{O}^{\bullet\bullet}] 2q\mu_v \quad (11.5)$$

Optimized levels of σ_i obviously require a combination of high charge carrier density and mobility.

Classically, high charge carrier densities have been induced in solids by substituting lower valent cations for the host cations [11.2]. Implicit in the requirement for high carrier densities are:

1. High solid solubility of the substituent with the lower valency,
2. Low association energies between the oxygen vacancy and dopant,
3. No long-range ordering of defects.

Additives which induce minimal strain tend to exhibit higher levels of solubility. The fluorite structure is the most well-known of these structures, with stabilized zirconia the best-known example. In this case, Y^{3+} substitutes for approximately 10% of Zr in $\text{Zr}_{1-x}\text{Y}_x\text{O}_{2-x/2}$, leading to $\sigma_i \approx 10^{-1} \text{ S/cm}$ at 1000°C and an activation energy of $\approx 1 \text{ eV}$. Other examples include CeO_2 [11.8], other fluorite-related structures such as the pyrochlores $\text{A}_2\text{B}_2\text{O}_7$ [11.9], and perovskites such as $\text{La}_{1-x}\text{Sr}_x\text{Ga}_{1-y}\text{Mg}_y\text{O}_{3-\delta}$ (LSGM) [11.10].

Since the dopant and vacancy are of opposite charge (for example, Y_{Zr}' and $\text{V}_\text{O}^{\bullet\bullet}$), they tend to associate. With cations being much less mobile than oxygen ions, this serves to trap the charge carrier. It is of interest to examine how the concentration of “free” mobile carriers

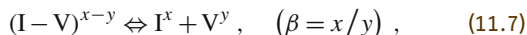
Table 11.1 Typical defect reactions

Defect reactions	Mass action relations	
$\text{MO} \rightleftharpoons \text{V}_\text{M}'' + \text{V}_\text{O}^{\bullet\bullet}$	$[\text{V}_\text{M}''][\text{V}_\text{O}^{\bullet\bullet}] = K_S(T)$	(1)
$\text{O}_\text{O} \rightleftharpoons \text{V}_\text{O}^{\bullet\bullet} + \text{O}_\text{i}'$	$[\text{V}_\text{O}^{\bullet\bullet}][\text{O}_\text{i}'] = K_F(T)$	(2)
$\text{O}_\text{O} \rightleftharpoons \text{V}_\text{O}^{\bullet\bullet} + 2\text{e}' + 1/2\text{O}_2$	$[\text{V}_\text{O}^{\bullet\bullet}]n^2 = K_R(T)P_{\text{O}_2}^{-1/2}$	(3)
$0 \rightleftharpoons \text{e}' + \text{h}^\bullet$	$np = K_e(T)$	(4)
$\text{N}_2\text{O}_3(\text{MO}_2) \rightleftharpoons 2\text{N}_\text{M}' + 3\text{O}_\text{O} + \text{V}_\text{O}^{\bullet\bullet}$	$[\text{N}_\text{M}']^2 \cdot [\text{V}_\text{O}^{\bullet\bullet}] / a_{\text{N}_2\text{O}_3} = K_N(T)$	(5)

depends on the dopant concentration and the association energy. Consider the neutrality relation representing vacancy compensation of acceptor impurities by

$$N_V = \beta N_I, \quad (11.6)$$

where N_V and N_I are the vacancy and impurity densities while β reflects the relative charges of the two species and normally takes on values of 1 (for A_M'') and $\frac{1}{2}$ (for A_M'). The association reaction is given by



where x and y are the relative charges of the impurity and vacancy, respectively. The corresponding mass action relation is then

$$N_I N_V / N_{\text{Dim}} = K_A^\circ \exp(-\Delta H_A / kT) \quad (11.8)$$

where N_{Dim} is the concentration of dimers and N_I and N_V are the corresponding defects remaining outside the complexes. It is straightforward to show that for weak dissociation (low temperatures or high association energies) one obtains the following solutions:

$$\beta = 1: \quad N_V = \left(N_I K_A^\circ\right)^{\frac{1}{2}} \exp(-\Delta H_A / 2kT), \quad (11.9)$$

$$\beta < 1: \quad N_V = \left(\frac{1-\beta}{\beta}\right) K_A^\circ \exp(-\Delta H_A / kT). \quad (11.10)$$

The solution for condition $\beta=1$ is the more familiar one. As in semiconductor physics [11.11], the number of free electrons or holes is proportional to the square root of the dopant density at reduced temperature, and it exhibits an Arrhenius dependence with activation energy that is

equal to one half of the association or ionization energy. The solution for condition $\beta < 1$ is more unusual. Here one predicts that N_V is independent of dopant density! Also, the activation energy is predicted to be equal to the association energy.

At sufficiently high temperatures or low association energies, essentially all of the dimers are dissociated and

$$N_V = \beta N_I = \beta N_I (\text{total}). \quad (11.11)$$

In general, therefore, two energies contribute to ionic conduction: a defect energy, E_D (which may either be related to the Frenkel or Schottky formation energy, or to a dissociation energy), and a migration energy E_m . The value of E in (11.3) therefore takes on different values in three characteristic temperature regimes. These include:

1. $E = E_m + E_A/2$: extrinsic associated regime at low T ,
2. $E = E_m$: extrinsic fully dissociated regime at intermediate T ,
3. $E = E_m + E_F/2$: intrinsic defect regime at elevated T (for instance, for Frenkel equilibrium).

For optimized ionic conduction to exist, two criteria must be satisfied simultaneously. First the term c in (11.3) must approach $1/2$. This corresponds to nearly all of the ions on a given sublattice being mobile. Second, the crystal structure must be arranged so as to enable easy motion of ions from one equivalent site to the next. This is reflected in exceptionally low values for the migration energy E_m . In the next section we discuss the conditions under which these criteria are satisfied.

11.2 Fast Ion Conduction

A number of routes leading to exceptionally high ion carrier densities in solids have been identified over the last few decades. These are subdivided into two major categories below (structurally disordered solids and highly defective solids). An important new development in recent years is the focus on the role of interfaces in creating ionic disorder localized in the vicinity of the boundaries. For nanosized structures, these disordered regimes may represent a large fraction of the overall volume of the material. Whatever the source of the enhanced ionic conductivity, such solids are commonly designated as *fast ion conductors* (FIC).

11.2.1 Structurally Disordered Crystalline Solids

In contrast to the idealized picture of crystal structures, many solids exist in which a sublattice of sites is only partially occupied. Strook [11.12, 13] already came to this conclusion in the 1930s in relation to the Ag sublattice in the high-temperature form (α -phase) of AgI. More recent neutron diffraction studies [11.14] differ with regard to the number of equivalent Ag sites. The special feature of partial occupancy of sites is nevertheless sustained.

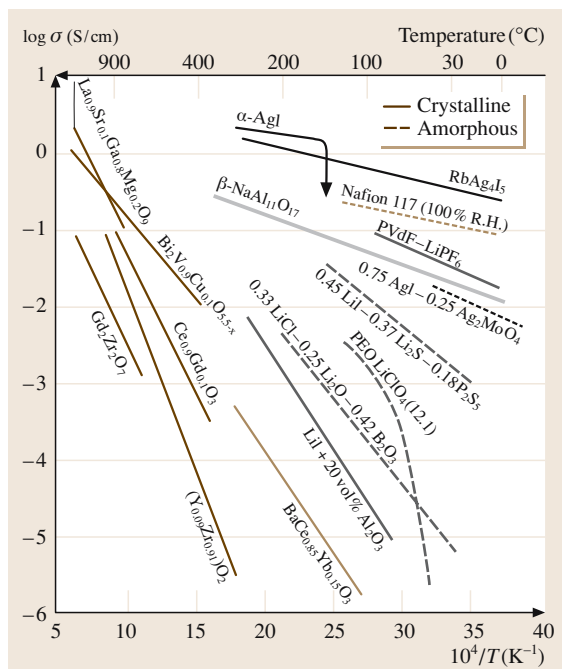


Fig. 11.2 The temperature dependences of representative FICs, including cation and anion conductors, crystalline and amorphous conductors, and inorganic and organic conductors. After [11.3]

Other notable systems characterized by sublattice disorder include Nasicon ($\text{Na}_3\text{Zr}_2\text{PSi}_2\text{O}_{12}$), sodium beta alumina ($1.2 \text{ Na}_2\text{O} - 0.11 \text{ Al}_2\text{O}_3$) and LiAlSiO_4 ,

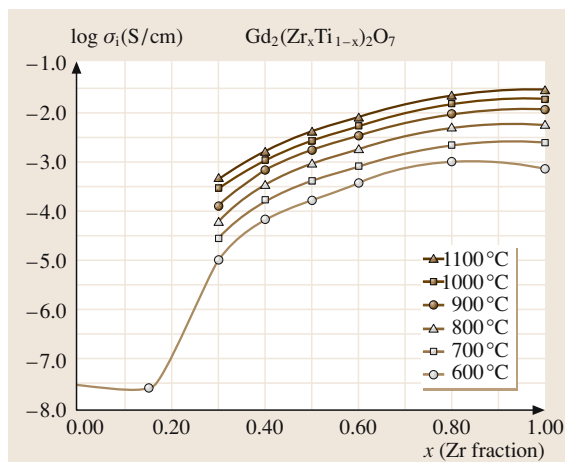


Fig. 11.3 Log ionic conductivity versus mole fraction Zr in $\text{Gd}_2(\text{Zr}_x\text{Ti}_{1-x})_2\text{O}_7$ at a series of temperatures. After [11.16]

which exhibit fast ion transport in three, two and one dimensions, respectively. Hundreds of other structurally disordered conductors may be found listed in review articles on the subject [11.2, 15]. Figure 11.2 illustrates the $\log \sigma - 1/T$ relations for representative FICs, while Table 11.2 summarizes data on representative materials in tabular form.

Similarities between FICs and liquid electrolytes are often noted; the most important of these is that the disordered sublattice in the solid resembles the disordered nature of ions in a liquid. For this reason one often hears the term *lattice melting* used to describe phase transitions in solids which relate to the conversion of a conventional ionic conductor to a FIC (such as β to α transition in AgI at approximately 150°C). Nevertheless, most investigators now believe that transport in FICs occurs via correlated jumps between well-defined sites rather than the liquid-like motion characteristic of aqueous or molten salt electrolytes.

The major structural characteristics of FICs include: (a) a highly ordered, immobile or framework sublattice providing continuous open channels for ion transport, and (b) a mobile carrier sublattice which supports a random distribution of carriers over an excess number of equipotential sites. FICs exhibit framework sublattices that minimize strain, electrostatic and polarization contributions to the migration energy while offering high carrier concentrations within the mobile carrier sublattice. Given the high concentration of carriers, correlation effects between carriers must be taken into account. Calculations by Wang et al. [11.17], for example, have demonstrated that cooperative motions of ions can lead to significantly lower calculated migration energies than those based on consideration of isolated jumps alone. Although no precise criterion now exists for categorizing FICs, they normally exhibit unusually high ionic conductivities ($\sigma > 10^{-2} \text{ S/cm}$), well below their melting points, and generally low activation energies (commonly $E \approx 0.05 - 0.5 \text{ eV}$, but they can be as high as $\approx 1.0 \text{ eV}$).

Few intrinsically disordered oxygen ion conducting FICs are known. The pyrochlores, with general formula $\text{A}_2^{3+}\text{B}_2^{4+}\text{O}_7$, represent a particularly interesting system, given that the degree of disorder can be varied almost continuously from low to high values within select solid-solution systems. The pyrochlore crystal structure is a superstructure of the defect fluorite lattice with twice the lattice parameter and one out of eight oxygens missing. This can be viewed as resulting from a need to maintain charge neutrality after substituting trivalent ions for 50% of the quadrivalent ions in the fluorite structure, as in $\text{Gd}_2\text{Zr}_2\text{O}_7$. Although

Table 11.2 Representative solid electrolytes and mixed conductors: mobile ions, electrical properties and applications

Material	Mobile ion	Properties	Remarks and applications
Zr _{0.85} Y _{0.15} O _{2-x} [11.20]	O ²⁻	$\sigma_{O^{2-}}$ (1000 °C) = 0.12 S/cm $E = 0.8$ eV	Material of choice in auto exhaust sensors; prime solid electrolyte candidate for SOFCs
Ce _{0.95} Y _{0.05} O _{2-x} [11.21]	O ²⁻	$\sigma_{O^{2-}}$ (1000 °C) = 0.15 S/cm $E = 0.76$ eV	Semiconducting at low PO ₂ ; prime candidate solid electrolyte for intermediate temperature SOFCs
Ce _{0.1} Gd _{0.9} O _{2-x} [11.22]	O ²⁻	$\sigma_{O^{2-}}$ (800 °C) = 0.10 S/cm $E = 0.56$ eV	
La _{0.8} Sr _{0.2} Ga _{0.8} Mg _{0.2} O _{3-x} [11.23]	O ²⁻	$\sigma_{O^{2-}}$ (750 °C) = 0.35 S/cm $E = 0.55$ eV	Readily converts to mixed conductor by addition of transition metals; candidate solid electrolyte for intermediate temperature SOFCs
Gd _{1.8} Ca _{1.2} Ti ₂ O _{7-x} [11.24]	O ²⁻	$\sigma_{O^{2-}}$ (1000 °C) = 0.05 S/cm $E = 0.67$ eV	The related pyrochlore, Gd ₂ Zr ₂ O ₇ , is an intrinsic FIC
Bi ₂ V _{0.9} Cu _{0.1} O _{5.5-x} [11.25]	O ²⁻	$\sigma_{O^{2-}}$ (700 °C) = 0.15 S/cm $E = 0.47$ eV	Mixed ionic–electronic conductor, $t_{O^{2-}} \approx 0.9$ at 900 K; of interest as a permeation membrane
La _{0.8} Sr _{0.2} MnO _{3+x} [11.26]	O ²⁻	$\sigma_{O^{2-}}$ (1000 °C) $\approx 3 \times 10^{-7}$ S/cm $E = 2.81$ eV	This material is largely an electronic conductor with σ_e (1000 °C) ≈ 100 S/cm; prime candidate as a cathode in SOFC .
BaCe _{0.85} Yb _{0.15} O _{3-x} [11.27]	H ⁺	σ_{H^+} (300 °C) = 7×10^{-4} S/cm $E = 0.54$ eV	Of interest in SOFC based on protonic conduction
Nafion [11.28]	H ⁺	σ_{H^+} (75 °C) $\approx 10^{-2}$ S/cm (at 20% relative humidity)	Organic; prime candidate for low-temperature solid state fuel cell based on protonic conduction
α – AgI [11.29]	Ag ⁺	σ_{Ag^+} (200 °C) = 1.6 S/cm $E = 0.1$ eV	Phase transition at 146 °C; first recognized fast ion conductor
RbAg ₄ I ₅ [11.29]	Ag ⁺	σ_{Ag^+} (30 °C) = 0.3 S/cm $E = 0.09$ eV	One of the most conductive ionic conductors at room temperature
α – CuI [11.30]	Cu ⁺	σ_{Cu^+} (450 °C) = 10^{-1} S/cm $E = 0.15$ eV	Phase transition at 407 °C
Na β -alumina [11.31]	Na ⁺	σ_{Na^+} (300 °C) = 0.13 S/cm $E \approx 0.3$ eV	Stoichiometry varies between Al ₂ O ₃ /Na ₂ O = 5.3–8.5
60Li ₂ S – 40SiS ₂ [11.32]	Li ⁺	σ_{Li^+} (25 °C) = 5×10^{-4} S/cm $E = 0.25$ eV	Amorphous
poly(vinylidene fluoride) (PVdF) – propylene carbonate (PC) – Li salt (LiX = LiSO ₃ CF ₃ LiPF ₆ or LiN(SO ₂ CF ₃) ₂) [11.33]	Li ⁺	σ_{Li^+} (20 °C) $\approx 10^{-3}$ S/cm	Organic conductor; of interest for lithium batteries
β -PbF ₂ [11.34]	F ⁻	σ_{F^-} (100 °C) = 10^{-4} S/cm $E = 0.48$ eV	Basis of a variety of gas sensors

oxygen vacancies occur at random throughout the anion sublattice in an ideal defect fluorite (such as **YSZ**), they are ordered onto particular sites in the pyrochlore structure. Thus, one properly views these as empty interstitial oxygen sites rather than oxygen vacancies. As a consequence, nearly ideal pyrochlore oxides, such as Gd₂Ti₂O₇, are ionic insulators [11.18]. Figure 11.3 illustrates the large increases in ionic conductivity induced

by systematically substituting zirconium for titanium. *Moon and Tuller* [11.16] explain this on the basis of increased A and B cation antisite disorder as the radius of the B ion approaches that of the A ion. Thus, as the cation environments of the oxygen ions becoming more homogeneous, exchange between regular and interstitial sites also becomes more favorable, leading to increased Frenkel disorder. This interpretation has been

confirmed by neutron diffraction studies on a closely related system [11.19].

Other important intrinsically disordered oxygen ion conductors are based on Bi_2O_3 . At 730°C [11.35], the low-temperature semiconducting modification transforms to the δ phase, which is accompanied by an oxide-ion conductivity jump of nearly three orders of magnitude. This is tied to the highly disordered fluorite-type structure, where a quarter of the oxygen sites are intrinsically empty, and to the high polarizability of the bismuth cation. *Takahashi* and *Iwahara* [11.36] succeeded in stabilizing the high-temperature δ phase to well below the transition temperature by doping with various oxides, including rare-earth oxides such as Y_2O_3 . High oxide ion conductivity was also discovered above 570°C in the Aurivillius-type γ phase of $\text{Bi}_4\text{V}_2\text{O}_{11}$ [11.37, 38], where one quarter of the oxygen sites coordinating V^{5+} are empty. Partial substitution of vanadium by lower valence cations, such as copper, nickel or cobalt, led to a new family of so-called BIMEVOX compounds [11.39], with a remarkably high oxygen ion conductivity at moderate temperatures. The copper-substituted compound has an oxide ion conductivity above 200°C which is ≈ 2 orders of magnitude higher than other oxide ion conductors. The bismuth-based electrolytes unfortunately suffer from instability under reducing conditions, which is a limitation for some applications, such as in the solid oxide fuel cells discussed below.

11.2.2 Amorphous Solids

One of the oft-mentioned criteria for FIC in solids (see before) is the existence of a highly ordered framework which provides channels for the ready motion of ions in the complementary, disordered sublattice. Reports of FIC in inorganic glasses [11.40] raised serious doubts concerning the relevance of this feature. The amorphous state, viewed as being liquid-like, is known to lack long-range order, with short-range order typically extending to, at most, a few atom spacings. Although highly oriented channels may be helpful in FIC, they are not essential, as demonstrated by the existence of FIC in glasses.

Fast ionic conductivity is observed in many glasses containing smaller cations with mole fractions greater than about 0.20, such as silver, copper, lithium and sodium [11.41, 42]. These glasses typically contain one or more network formers (such as SiO_2 , B_2O_3 , P_2O_5 or GeS_2), network modifiers (such as Ag_2O , Li_2O , Cu_2O or Ag_2S), and dopant compounds, largely halides (such as

AgI , CuI and LiCl). The network structure and therefore its physical and chemical properties can be substantially modified by addition of the modifier. Dopant salts, on the other hand, do not strongly interact with the network, but *dissolve* into the interstices of the glass structure. A number of phenomenological trends have been noted including: ion conduction increases (1) in the order K , Na , Li , Ag , (2) with increasing modifier concentration, (3) with halide additions in the order Cl , Br , I , (4) in sulfide versus oxide glasses, and (5) in correlation with decreasing density and glass transition temperature.

Some authors speculate that ionic transport in glasses is enhanced upon addition of the halide anions by lowering the association energy between the mobile charge carriers and the network and thereby increasing the free carrier density [11.43]. An alternate model attributes the increased conductivity to major changes induced in the glass structure by the additives, as reflected in changes in glass transition temperature T_g and the density ρ . In this latter model, a large fraction of the carriers are already assumed to be unassociated and free to move, but with increased ionic mobility driven by structural changes. Here [11.44–47], the predominant influence of the halide addition is believed to impact the strain component of the migration energy.

Another important class of amorphous fast ion conductors is those based on organic or polymer electrolytes, which (analogous to the inorganic systems) are composed of a backbone polymer and a salt complex in which the counter-ion is covalently bound to the backbone. A classic example is the one based on polypropylene oxide $(\text{CH}_2\text{CH}(\text{CH}_3)\text{O})_n$ (PPO) complexed with LiCF_3SO_3 to form $\text{PPO}_n \cdot \text{LiCF}_3\text{SO}_3$. Upon forming the complex, the Li ion conductivity increases by as much as a factor of $\approx 10^5$ [11.48]. In contrast to the inorganic glasses, which exhibit an Arrhenius temperature dependence, however, these polymers follow a curved dependence best expressed by

$$\sigma = \sigma_0 \exp\left(-\frac{B}{T - T_0}\right), \quad (11.12)$$

where T_0 is the glass transition temperature. This suggests a coupling between transport and network relaxation, a situation more closely coupled to transport in a liquid than in a solid, albeit a highly viscous liquid. Polymer electrolytes are now materials of choice for Li batteries and proton-based solid electrolytes given their attractive mechanical properties (ability to relax elastically upon stresses induced by volume changes related to charge/discharge of adjacent electrodes) and ease of processing [11.49].

11.2.3 Heavily Doped Defective Solids

Anomalously high concentrations of ionic carriers may also be induced in intrinsically insulating solids. In the following we briefly discuss two approaches for generating such *highly defective solids*.

We already know that ionic defect densities may be greatly enhanced above intrinsic levels by doping with altermultivalent impurities. However, the solubility limit of such impurities is often limited to only tens or hundreds of ppm. This corresponds to roughly 10^{17} – 10^{18} defects/cm³, a value 10^3 – 10^4 times smaller than in typical FICs. Compounds do exist, however, in which the solubility limit is extensive, reaching the 10–20% level even at reduced temperatures. Perhaps the most familiar example of such a system is stabilized zirconia, which due to its wide solid solubility with cations of lower valency such as Ca²⁺ and Y³⁺, exhibits exceptionally high oxygen ion conductivity ($\sigma \approx 10^{-1}$ S/cm) at temperatures approaching 1000 °C.

As discussed above, high carrier densities must be coupled with high ion mobilities in order to attain high magnitudes of ionic conduction. The cubic fluorite structure, exhibited by stabilized zirconia (ZrO₂) and ceria (CeO₂), for example, supports high oxygen ion mobility due to the low four-fold coordination of cations around the oxygens, coupled with the interconnected nature of the face-shared polyhedra which surround the oxygen sites. Migration energies as low as ≈ 0.6 eV are reported for oxygen vacancy motion in ceria-based solid solutions [11.50]. High fluorine ion mobility is also observed in fluorite CaF₂ and related crystal systems.

More recently Ishihara demonstrated that very high oxygen conductivity can be achieved in the perovskite LaGaO₃ by acceptor doping on both the La and Ga sites [11.51]. The solid solution (La_{1-x}Sr_x)(Ga_{1-y}Mg_y)O₃ exhibits ionic conductivity levels above that of ZrO₂ and CeO₂, for example 3×10^{-1} S/cm at 850 °C. Perovskites also support some of the highest proton conductivities at elevated temperatures. The most popular of these are ABO₃-type compounds with A = Ba, Sr, and B = Ce or Zr. Upon acceptor doping, as in SrCe_{0.95}Yb_{0.05}O₃, oxygen vacancies are generated as in the gallate above. However, in the presence of moisture, water is adsorbed and protons are generated [11.52]:



Given the high proton mobility, this is sufficient to induce large proton conductivity. Perovskite-related structures with the general formula A₃B'B''O₉ also

exhibit high protonic conductivity [11.53]. Atomistic calculations simulating proton diffusion in numerous perovskite-type oxides are reported by the group of Catlow [11.54].

Ionic conductivities do not generally increase linearly with foreign atom additions. At the levels of defects being discussed here, defect–defect interactions become important, generally leading to defect ordering. This results in a maximum in ionic conductivity at some level of doping that depends on the particular system being investigated. Nowick and coworkers [11.55, 56] have demonstrated, in a series of studies, that the deviations from ideality are caused initially by composition-dependent activation energies rather than pre-exponentials (11.9), a feature also observed in a number of FIC glasses.

The formation of ionic defects that accompany excursions in composition away from stoichiometry due to redox reactions (Table 11.1, reaction 3) may also be large. CeO₂, for example, may be readily reduced to CeO_{1.8} at 1000 °C [11.57], resulting in oxygen vacancy concentrations of 5×10^{21} cm⁻³ ($c = 0.9$). It should be noted, however, that comparable concentrations of electrons are also formed during such stoichiometry excursions.

11.2.4 Interfacial Ionic Conduction and Nanostructural Effects

Interfaces can significantly modify the ionic conductivities of polycrystalline or composite materials and thin films. Modified levels of ionic conductivity near interfaces may result from space-charge regions formed near interfaces to compensate for charged defects and impurities segregated to surfaces, grain and phase boundaries. Grain boundaries, for example, serve as source and sink for impurities and point defects and thus often take on a net negative or positive charge relative to the grains. To maintain overall charge neutrality, a space charge of opposite charge forms in the grains adjacent to the grain boundaries with a width related to the Debye length L_D given by

$$L_D = (\epsilon_r \epsilon_0 k / T q^2 n_b)^{1/2} \quad (11.14)$$

in which n_b is the majority charge carrier concentration within the grain, $\epsilon_r \epsilon_0$ the dielectric constant, k_B the Boltzmann constant, T the temperature and q is the electron charge. Depending on the sign of the charge at the interface, a depletion or accumulation of mobile ions in the vicinity of the boundary will form. Liang provided

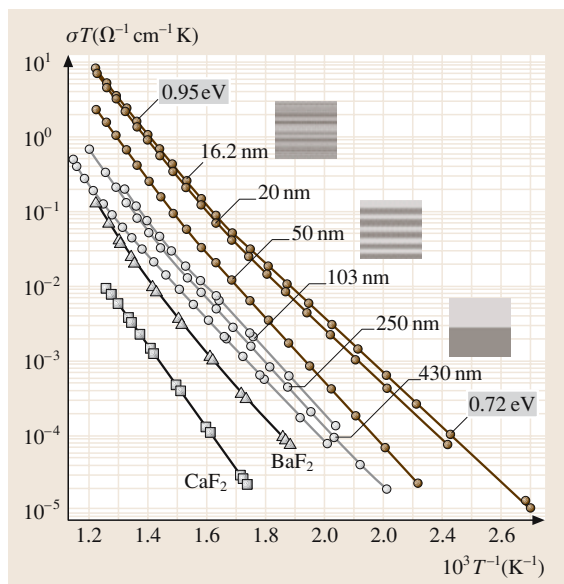


Fig. 11.4 Parallel ionic conductivity of $\text{CaF}_2/\text{BaF}_2$ nanometer-scale, artificially modulated heterolayers, with various periods and interfacial densities in the 430 to 16 nm range. After [11.59]

one of the first demonstrations of enhancement in the $\text{LiI} : \text{Al}_2\text{O}_3$ system [11.58].

The defect concentration profile in the space-charge region can be expressed as [11.60]

$$c_i/c_i^\infty = \exp \left[-q_i(\phi - \phi^\infty)/k_B T \right]. \quad (11.15)$$

The bulk concentration (c_i^∞) is a function of temperature, chemical potential and doping. The local concentration in the space-charge region (c_i) depends

on the difference between the bulk and the local electrical potential (ϕ^∞ and ϕ). For positive values of ϕ , the concentrations of all negative defects are increased by the exponential factor, while those of the positive defects are decreased by the same factor and vice versa for negative values.

Films and/or polycrystalline materials with very small lateral dimensions can be expected to exhibit particularly strong space-charge effects on ionic conduction. This follows from the fact that the space-charge width approaches the dimensions of the film or grain. In this case, the space-charge regions overlap, and the defect densities no longer reach bulk values, even at the center of the particles [11.60]. In the limit of very small grains, local charge neutrality is not satisfied anywhere, and a full depletion (or accumulation) of charge carriers can occur with major consequences for ionic and electronic conductivity. Strong nanoscale effects on ionic and mixed ionic conductivity have been demonstrated for artificially modulated heterolayers of the solid ionic conductors $\text{CaF}_2/\text{BaF}_2$ [11.59] and nanocrystalline CeO_2 [11.61]. Figure 11.4 illustrates the orders of magnitude increase in fluorine ion conductivity possible with space-charge accumulation of mobile carriers in nanoscale $\text{CaF}_2/\text{BaF}_2$ multilayers.

11.3 Mixed Ionic–Electronic Conduction

11.3.1 Defect Equilibria

Deviations from stoichiometry in the direction of oxygen excess (MO_{1+x}) or deficiency (MO_{1-x}) form defect states that act identically in every way to impurity-related acceptor or donor states, respectively. In general, the electrical behavior of solids depends on defects formed in response to both impurities and deviations from stoichiometry. At or near stoichiometry, impurities predominate, while under strongly reducing or oxidizing conditions, defects associated with deviations from stoichiometry often take control. To characterize the electrical response of a metal oxide to temperature and atmosphere excursions, a series of simultaneous reactions of the form represented by (Table 11.1, reactions

1–5) must be considered. Furthermore, a representative electroneutrality equation for the case considered in Table 11.1 would be:

$$2[\text{O}_i''] + [\text{N}_M'] + n = 2[\text{V}_\text{O}^{\bullet\bullet}] + p. \quad (11.16)$$

Note that: (1) intrinsic Frenkel disorder is assumed to predominate, so that (Table 11.1, reaction 1) may be ignored in subsequent discussions; (2) aN_2O_3 is often assumed to be sufficiently low that all of N goes into solid solution.

A piecewise solution to such problems is commonly attempted by sequentially choosing conditions for which only one term on either side of (11.16) need be considered. The region corresponding to mixed ionic conductivity is where the predominant charge carrier is

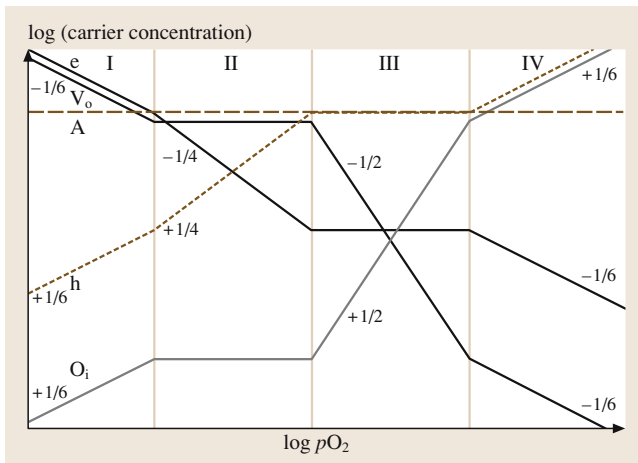


Fig. 11.5 Defect diagram for acceptor-doped oxide. After [11.62]

an ion. In acceptor-doped material (A' being a generic acceptor), this corresponds to the condition (Region II in Fig. 11.5) for which (11.16) may be simplified to read

$$[N'_M] = [A'_c] = 2[V_O^{\bullet\bullet}]. \quad (11.17)$$

Combining this with (Table 11.1, reaction 3) one obtains

$$n = \left[2K_R(T) / [N'_M] \right]^{1/2} P_{O_2}^{-1/4} \quad (11.18)$$

and from (Table 11.1, reactions 2 and 4),

$$p = K_e(T) \left(2K_R(T) / [N'_M] \right)^{-1/2} P_{O_2}^{1/4}, \quad (11.19)$$

$$[O_i''] = 2[A'_c]^{-1} K_F(T). \quad (11.20)$$

Note that, in this defect regime, the ionic defects are pO_2 -independent while the electronic species exhibit a $pO_2^{\pm 1/4}$ dependence. One obtains predictions for the corresponding dependencies of the partial conductivities of each of these charged species by multiplying carrier concentration by the respective charge and mobility. Experimentally, one normally observes the same pO_2 dependence of the partial conductivity as that predicted for the defect concentration, demonstrating that the mobility is pO_2 -independent. One then uses the predicted pO_2 dependencies of the partial conductivities to deconvolute the ionic and electronic contributions to the electrical conductivity as discussed below.

The three other defect regimes most likely to occur, beginning at low P_{O_2} and moving on to increasing P_{O_2} , are depicted in Fig. 11.5 and include $n = 2[V_O^{\bullet\bullet}]$ (Region I), $p = [N'_M]$ (Region III) and $p = 2[O_i'']$ (Region IV).

In the case where μ_n, μ_p are sufficiently greater than $\mu(V_O^{\bullet\bullet})$, then even in the defect regime where $V_O^{\bullet\bullet}$ is the predominant defect (so that $2[V_O^{\bullet\bullet}] = [N'_M]$), the total conductivity remains electronic. When the carrier mobility inequality is not nearly so pronounced, so that at the pO_2 at which electronic defects are at a minimum ($n = p$), conduction is predominantly ionic. Under these circumstances the oxide acts as a solid electrolyte, and in this regime of temperature and pO_2 , one designates this as the *electrolytic domain*. Aside from the electrolytic domain, the neighboring zones on either side are designated as mixed zones within which both ionic and electronic conductivities are of comparable magnitude.

11.3.2 Electrolytic Domain Boundaries

In applications where solid electrolytes are to be utilized, it is essential to know a priori under which conditions the material is likely to exhibit largely electrolytic characteristics. Expressions for the *electrolytic domain boundaries* can be obtained by first writing down general expressions for the partial conductivities (11.17–11.19):

$$\sigma_i = \sigma_i^\circ \exp(-E_i/kT), \quad (11.21)$$

$$\sigma_p = \sigma_p^\circ P_{O_2}^{+1/4} \exp(-E_p/kT), \quad (11.22)$$

$$\sigma_n = \sigma_n^\circ P_{O_2}^{-1/4} \exp(-E_n/kT). \quad (11.23)$$

One commonly defines the electrolytic domain boundary as that condition of T and pO_2 for which the ionic conductivity drops to 0.5. Under reducing conditions, this pO_2 is designated by P_n and under oxidizing conditions by P_p . Consequently, one equates σ_i and σ_n or σ_i and σ_p to solve for P_n and P_p , respectively. These are given by

$$\ln P_n = \frac{-4(E_n - E_i)}{k} \frac{1}{T} + 4 \ln \left(\frac{\sigma_n^\circ}{\sigma_i^\circ} \right), \quad (11.24)$$

$$\ln P_p = \frac{-4(E_i - E_p)}{k} \frac{1}{T} + 4 \ln \left(\frac{\sigma_i^\circ}{\sigma_p^\circ} \right). \quad (11.25)$$

Note that since the mobilities of vacancies in such oxides have been found to be much greater than those of interstitials [11.16], we ignore the latter's contributions. The domain boundaries for stabilized zirconia are shown plotted in Fig. 11.6 [11.63]. Note that, as commonly observed, the electrolytic domain shrinks with increasing temperature due to the fact that E_n and E_p are typically greater in magnitude than E_i .

11.4 Applications

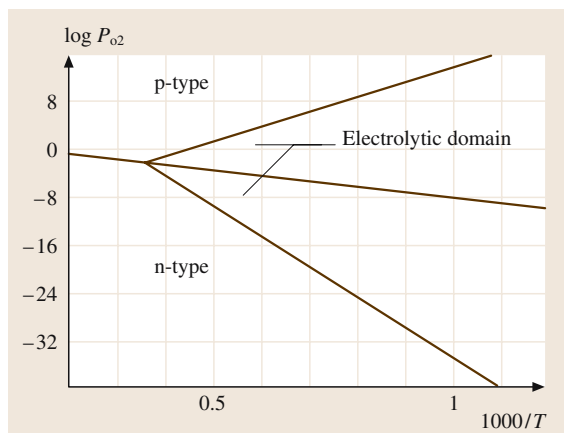


Fig. 11.6 Domain boundaries of stabilized zirconia as projected onto the $\log P_{O_2} - 1000/T$ plane [11.63]

Fast ionic conducting ceramics and MIECs are finding extensive application in various solid state electrochemical devices. Some of these include fuel cells and electrolyzers [11.2, 64–66], high energy density Li batteries [11.67], electrochromic windows [11.68] and auto exhaust sensors [11.69].

11.4.1 Sensors

The monitoring of our environment has become essential for effective emissions control. Likewise, monitoring of chemical processes in real time enables closer quality control of products. Electrochemical sensors transform a chemical signal into an electrical signal, which is easy to measure, monitor and process [11.70]. Ionic and mixed conducting solids are basic materials for

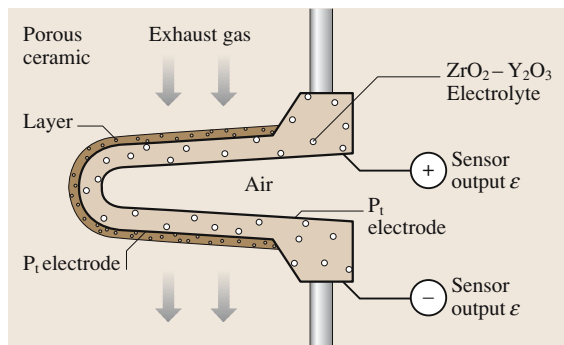


Fig. 11.7 Schematic of auto exhaust sensor based on the Nernst equation. After [11.64]

this development, because they can be easily miniaturized, for instance in thin film form, and they can often be operated at elevated temperatures or in an aggressive environment. The two major types of electrochemical sensors, potentiometric and amperometric, are summarized below.

Potentiometric Sensors

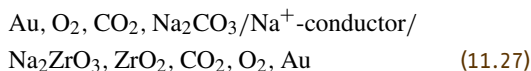
In a potentiometric gas sensor, the concentration or partial pressure of a species is determined by measuring the emf of a solid electrolyte concentration cell. The most successful commercial sensor is the oxygen sensor [11.64], which uses stabilized zirconia as the solid oxygen ion electrolyte (Fig. 11.7). The emf of the cell, $O_{2,ref}$, Pt / YSZ / Pt, O_2 , can be written according to the Nernst equation:

$$E = t_i(kT/4q) \ln[P(O_2)/P(O_2)_{ref}] . \quad (11.26)$$

$P(O_2)_{ref}$ is the oxygen partial pressure of the reference gas, generally air, and t_i is the ionic transference number. For proper operation, t_i needs to be kept very close to unity. All other terms have their common meanings.

The zirconia auto exhaust sensor monitors the air-to-fuel ratio, which is maintained within close limits for optimum operating efficiency of the three-way exhaust catalyst that serves to reduce the amount of pollutants, including unburned hydrocarbons, CO and NO_x . Such sensors are designed to provide response times on the order of tens of milliseconds. A voltage near to zero corresponds to an oxygen-rich “lean” mixture, and a voltage near to 1 V to an oxygen-poor “rich” mixture. The fuel injector of the engine is controlled via a closed-loop system. The zirconia oxygen sensor sees a wide range of exhaust temperatures, up to values as high as $\approx 900^\circ\text{C}$. Fortunately, the large step in voltage in going from lean to rich conditions can be easily detected at all temperatures. The molecular mechanisms operating at the electrolyte/platinum interface have been examined in detail [11.71].

Another example of a potentiometric sensor is the one reported by Maier et al., of the type [11.72]



This is used to monitor pCO_2 ; it eliminates the need for a gas-tight reference electrode. Here the two-phase reference electrode Na_2ZrO_3, ZrO_2 , which fixes the Na activity on the right side of the cell, is insensitive to

CO₂, while the Na₂CO₃ electrode on the left side can be sensed by the Na⁺-conductor, typically NASICON or β-alumina. Yamazoe and Miura have reviewed the possible different types of potentiometric sensors by using single or multicomponent auxiliary phases [11.73].

Amperometric Sensors

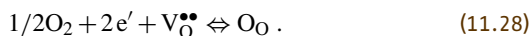
By applying a voltage across an electrolyte, it is possible to electrochemically pump chemical species from one chamber to the other. Amperometric sensors rely on limiting current due to diffusion or interfacial phenomena at the electrode, which are linearly dependent on the partial pressure of the gas constituent [11.74, 75]. These become particularly important in so-called lean burn engines. Here the partial pressure of oxygen does not strongly vary with the air-to-fuel ratio, in contrast to engines operating at or near the stoichiometric air-to-fuel ratio. Under these circumstances, sensors are needed which have a stronger than logarithmic sensitivity to oxygen partial pressure variations. Sensors based on this principle are also being developed to detect other gases including NO_x.

11.4.2 Solid Oxide Fuel Cells (SOFC)

Solid oxide fuel cells (SOFC) provide many advantages over traditional energy conversion systems, including high energy conversion efficiency, fuel flexibility (due to internal reforming), low levels of NO_x and SO_x emissions, versatile plant size and long lifetimes [11.76]. Quiet, vibration-free operation also eliminates noise associated with conventional power-generation systems. However, operation at elevated temperature is necessary given relatively low ionic conductivities and slow

electrode processes at temperatures below 800 °C. Recent progress with thinner electrolytes and advanced electrodes holds promise for reducing operating temperatures by as much as several hundred degrees.

The three major components of the elemental solid oxide fuel cell (SOFC) include the cathode, electrolyte and anode. While the solid electrolyte is selected so that it only conducts ions to ensure a Nernst open circuit potential that is as close to ideal as possible, the electrodes must support the reduction/oxidation reactions that occur at the electrolyte/electrode/gas interfaces. For example, when current is being drawn, the following reaction occurs at the cathode:



This reaction is accelerated if the cathode can provide both electrons, as in a typical current collector, as well as oxygen vacancies. An example of such a mixed conducting cathode is La_{1-x}Sr_xCoO₃ (LSCO), which has an electronic conductivity of > 100 S/cm and an oxygen ion conductivity of > 1 S/cm at temperatures above 800 °C [11.77]. Given its importance to performance, modeling of the electrode processes has also received a great deal of attention recently [11.78, 79]. Unfortunately, while exhibiting highly attractive mixed conducting properties, LSCO is unstable in contact with yttria-stabilized zirconia, the electrolyte of choice.

In an attempt to take advantage of LSCO's attractive features, there is growing interest in marrying this electrode with doped ceria electrolytes, such as Ce_{1-x}M_xO₂:M=Gd or Sm, for operation at reduced temperatures of 550–750 °C, given ceria's higher ionic conductivity (albeit higher mixed ionic electronic conductivity at the anode). Electronic conduction degrades solid electrolyte performance in several ways. Because electronic conduction serves as an alternate path for charged species through the electrolyte, it decreases the power that can be dissipated through the load. Further, the short circuiting factor also serves to allow permeation of gaseous species through the electrolyte, even under open circuit conditions (see the section on membranes below). These primary figures of merit are summarized in the context of a solid oxide fuel cell (SOFC) in Fig. 11.8. E represents the potential induced across the cell under open circuit conditions for a given P_{O_2} gradient, \hat{t}_i is the ionic transference number, E_N is the Nernst potential, R_{INT} , R_C , R_{SE} and R_A are the internal cell, cathode, solid electrolyte and anode resistances, respectively, J_{O_2} is the oxygen permeation flux, and L the thickness across which the P_{O_2} gradient is imposed. All other terms have their normal meanings. Fortunately, the

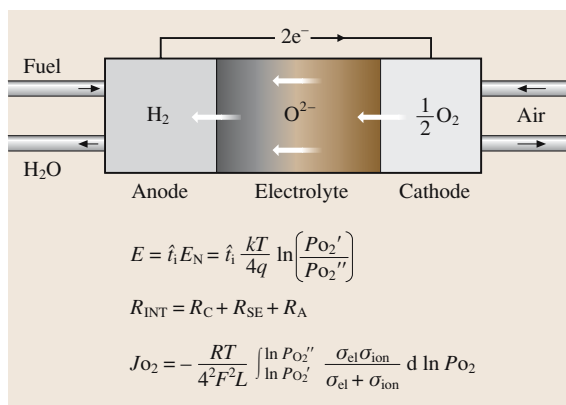


Fig. 11.8 Schematic of a solid oxide fuel cell and the primary figures of merit [11.66]

electronic conductivity in ceria electrolytes drops exponentially with decreasing temperature, and the overall power output exceeds that of zirconia-based systems at reduced temperatures, so it can be used with LSCO and other electrodes incompatible with YSZ.

The system $(\text{La}_{1-x}\text{Sr}_x)(\text{Ga}_{1-y}\text{Mg}_y)\text{O}_3$ (LSGM) was mentioned above as exhibiting one of the highest oxygen ion conductivities ($\approx 3 \times 10^{-1} \text{ S/cm}$ at 850°C) due to high levels of acceptor doping (Sr and Mg) on both the La and Ga sites. As a consequence, it is now being considered as one of several candidates for the electrolyte in solid oxide fuel cells. Experiments have shown that mixed ionic electronic conduction in a fuel cell electrode contributes to reduced overpotentials [11.66]. It has also been recognized that a single-phase *monolithic* fuel cell structure would benefit from the minimization of chemical and thermomechanical degradation [11.80]. Consequently, an electrode based on LSGM would satisfy all requirements. Long et al. proposed to add a transition metal in solution, which would introduce an additional 3d *conducting* band within the wide band-gap of the initially electronically insulating gallate [11.81]. As expected, as the Ni content in the system $\text{La}_{0.9}\text{Sr}_{0.1}\text{Ga}_{1-x}\text{Ni}_x\text{O}_3$ (LSGN) increased, the electronic conductivity increased, finally reaching $\approx 50 \text{ S/cm}$ without decreasing the already high ionic conductivity. Improved electrode performance was indeed observed with the LSGM/LSGN interface [11.82].

11.4.3 Membranes

Oxygen-permeable ceramic membranes are used for the separation of oxygen from air or for industrial-scale oxygen separation in the conversion of natural gas to syngas ($\text{CO} + \text{H}_2$) for example [11.83]. They are made from mixed conducting oxides in which ambipolar diffusion of ionic and electronic charge carriers in an oxygen potential gradient assures a high oxygen permeation flux through the membrane (see Fig. 11.8 for an expression for the permeation current). High oxygen permeation rates were obtained with the system $(\text{La}, \text{Sr})\text{MO}_{3-\delta}$ ($\text{M} = \text{Fe}, \text{Co}, \text{Cr}$) [11.77], but some deterioration over time was noticed. Research continues into this class of materials with regard to long-term ordering of defects, surface exchange kinetics, optimization of oxygen conduction and phase stability under steep oxygen activity gradients. One of the best materials developed to date is the BICUVOX compound, with composition $\text{Bi}_2\text{V}_{0.9}\text{Cu}_{0.1}\text{O}_{5.35}$, which shows a particularly large mixed conductivity that enables high oxygen permeation

rates at moderate temperature, such as 700 K, at high and intermediate oxygen partial pressures [11.38].

Mixed oxide ion and electronic conductivity is also observed in composites of a solid oxide ion electrolyte and a noble metal, if percolating pathways exist for each component. These mixed conducting oxide ceramic-metal composites (cermets), including Y-stabilized ZrO_2 with Pd [11.84], Sm-doped CeO_2 with Pd [11.85], and rare-earth doped Bi_2O_3 with Ag [11.86], have an appreciable oxygen permeation rate at elevated temperature without degradation and are considered attractive for industrial applications, although they are relatively expensive. Recent work by Takamura et al. [11.87] shows promising results based on ceramic/ceramic composites.

11.4.4 Batteries

Power storage requires high energy density batteries. The highest possible energy density is achieved using reactants with high free energies of reaction and low mass, such as lithium or sodium. This also requires that the solid electrolyte remains stable under highly reducing or highly oxidizing conditions. Major advantages of solid electrolytes over liquid electrolytes are the absence of leakage and container problems and, foremost, the possibility of miniaturization; for example using thin solid films.

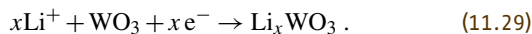
There is an increasing demand for microbatteries compatible with microelectronics technology, related to the development of laptop computers or portable telephones. This led to the development of high energy density and long life-cycle rechargeable lithium batteries, initially based on metallic lithium anodes. However, systems based on metallic lithium suffered from problems due to metal oxidation and poor rechargeability due to the formation of metallic dendrites. The alternative “rocking chair” concept, proposed in 1980 [11.88], based on two lithium insertion compounds Li_xWO_3 and Li_yTiS_2 , replaced the unstable lithium electrode, but was unable to provide sufficiently high energy densities. Improved rechargeable lithium ion batteries based instead on nongraphitic “hard” carbons as the lithium insertion anodes have since been developed [11.89]. This was followed by the successful association of hard carbon insertion anodes with the high-voltage LiCoO_2 insertion cathode. Due to the relatively high cost of Co, alternative systems based on other cathode materials, such as LiNiO_2 or LiMn_2O_4 , are currently under investigation [11.90]. Polymer (rather than inorganic) electrolytes are used in these applications (see

above). An overview of lithium batteries and polymer electrolytes can be found in books by *Julien* and *Nazri* [11.91] and *Gray* [11.92].

11.4.5 Electrochromic Windows

Electrochromic light transmission modulators – so-called “smart windows” that use solid ionic conductors – may play a significant role in energy-saving by regulating thermal insulation. In such a system, the window is maintained transparent in the visible and reflecting in the IR during the winter, allowing penetration of sunshine but blocking loss of interior heat. On the other hand, the window is rendered partially opaque during hot summer days, reducing the amount of radiation entering the building. The electrochromic elements [11.93], which color or bleach upon insertion/deinsertion of lithium or

hydrogen ions, are sandwiched between two transparent thin-film electrodes and are separated by a solid electrolyte. The transparent electrodes are generally indium tin oxide (ITO). Glass electrolytes appear to be promising choices. Research and development on tungsten trioxide-based electrochromic materials started in the 1970s. Ions fill empty tetrahedral sites in the WO₃ structure [11.94]:



Key requirements include: (1) compatible electrochromic and solid electrolyte thin film materials; (2) the ability to operate near ambient temperature, in the range -40 to $+120^\circ\text{C}$; (3) cycle reversibly many thousands of times a year for a lifetime of 20 years, and; (4) exhibit significant shifts in reflectivity with the degree of insertion/deinsertion.

11.5 Future Trends

A rapidly converging interest in thin film oxides has been developing in the microelectronics and solid state ionics communities. In the solid state ionics arena, the desire to reduce the operating temperature of solid oxide fuel cells (SOFC) has been stimulating a shift in emphasis from bulk to thin film electrolytes. Likewise, the trend in recent years has shifted away from bulk ceramics towards miniaturized smart sensor systems in which the sensor elements are integrated with electronics and various MEMS-based components, including microheaters, valves and membranes. Considering the continued drive towards ever smaller submicron lateral dimensions in MOSFET technology, it is likely that future efforts will be directed towards the construction of micro- and nanoscale ionic devices.

Specifically, one can envision the embedding of miniaturized thin film or SOFC structures as sensors or power sources together with microelectromechanical (MEM) components and other active electronics in the

same silicon wafer. By applying standard Si technology, such as thin film deposition and photolithography, one accesses methods for tailoring electrolyte and electrode geometry (thickness, active electrode area and triple phase boundary length) with exceptionally high dimensional reproducibility, while retaining the ability to scale to larger dimensions. Attention will need to be focused on the special challenges that the marriage between solid state ionics and electronics implies, including semiconductor-compatible processing, rapid temperature excursions, stress-induced property modifications and interfacial stability.

Finally, one will need to consider how the defect and transport properties of thin films may differ from their bulk counterparts, and how the silicon platform provides opportunities to examine such properties in an in situ manner, and thereby identify novel or distinctive properties associated with low-dimensional structures.

References

- | | |
|--|---|
| <p>11.1 H. L. Tuller: <i>J. Phys. Chem. Solids</i> 55, 1393–1404 (1994)</p> <p>11.2 P. Knauth, H. L. Tuller: <i>J. Am. Ceram. Soc</i> 85, 1654–1679 (2002)</p> | <p>11.3 H. L. Tuller: In: <i>Ceramic Materials for Electronics</i>, 3rd edn., ed. by R. C. Buchanan (Marcel Dekker, New York 2004) p. 87</p> <p>11.4 M. K. Paria, H. S. Maiti: <i>J. Mater. Sci.</i> 17, 3275 (1982)</p> |
|--|---|

- 11.5 J. B. Goodenough: In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) p. 393
- 11.6 F. A. Kröger: *The Chemistry of Imperfect Crystals*, 2nd edn. (North-Holland, Amsterdam 1974)
- 11.7 J. Hladik (ed.): *Physics of Electrolytes*, Vol. 1 (Academic, New York 1972)
- 11.8 B. C. H. Steele: *J. Power Sources* **49**, 1–14 (1994)
- 11.9 S. Kramer, M. Spears, H. L. Tuller: *Solid State Ionics* **72**, 59–66 (1994)
- 11.10 T. Ishihara, H. Matsuda, Y. Takita: *J. Am. Ceram. Soc.* **116**, 3801–3803 (1994)
- 11.11 S. Wang: *Fundamentals of Semiconductor Theory and Device Physics* (Prentice Hall, Englewood Cliffs 1989)
- 11.12 L. W. Strock: *Z. Phys. Chem.* **B25**, 441 (1934)
- 11.13 L. W. Strock: *Z. Phys. Chem.* **B25**, 132 (1936)
- 11.14 R. Cava, B. J. Wuensch: *Solid State Commun.* **24**, 411–416 (1977)
- 11.15 T. Kudo: In: *The CRC Handbook of Solid State Electrochemistry*, ed. by P. J. Gellings, H. J. M. Bouwmeester (CRC, Boca Raton 1997) p. 195
- 11.16 P. K. Moon, H. L. Tuller: *Solid State Ionics* **2** 8–3 0, 470–474 (1988)
- 11.17 J. Wang, M. Kaffari, D. Choi: *J. Chem. Phys.* **63**, 772 (1975)
- 11.18 S. A. Kramer, H. L. Tuller: *Solid State Ionics* **8**, 2, 15 (1995)
- 11.19 B. J. Wuensch, K. W. Eberman, C. Heremans, E. M. Ku, P. Onnerud, S. M. Haile, J. K. Stalick, J. D. Jorgensen: *Solid State Ionics* **12**(9), 111–133 (2000)
- 11.20 R. M. Dell, A. Hooper: In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) pp. 291–312
- 11.21 H. L. Tuller, A. S. Nowick: *J. Electrochem. Soc.* **122**, 255–259 (1975)
- 11.22 B. C. H. Steele: *J. Mater. Sci.* **36**, 1053–1068 (2001)
- 11.23 T. Ishihara, T. Shibayama, M. Honda, H. Nishiguchi, Y. Takita: *J. Electrochem. Soc.* **147**, 1332–1337 (2000)
- 11.24 S. Kramer, H. L. Tuller: *Solid State Ionics* **82**, 15–23 (1995)
- 11.25 A. A. Yaremchenko, V. V. Kharton, E. N. Naumovich, F. M. B. Marques: *J. Electroceram.* **4**, 233–242 (2000)
- 11.26 R. A. De Souza, J. A. Kilner, J. F. Walker: *Mater. Lett.* **43**, 43–52 (2000)
- 11.27 J. Wu, L. P. Li, W. T. P. Espinosa, S. M. Haile: *J. Mater. Res.* **19**, 2366 (2004)
- 11.28 F. Damay, L. C. Klein: *Solid State Ionics* **162–163**, 261–267 (2003)
- 11.29 J. S. Kasper: In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) pp. 217–235
- 11.30 T. Matsui, J. B. Wagner Jr.: In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) pp. 237–252
- 11.31 T. Takahashi: In: *Superionic Solids and Solid Electrolytes: Recent Trends*, ed. by A. L. Laskar, S. Chandra (Academic, San Diego 1989) pp. 1–41
- 11.32 F. A. Fusco, H. L. Tuller: In: *Superionic Solids and Solid Electrolytes: Recent Trends*, ed. by A. L. Laskar, S. Chandra (Academic, San Diego 1989) pp. 43–110
- 11.33 Z. Jiang, B. Carroll, K. M. Abraham: *Electrochim. Acta* **42**, 2667 (1997)
- 11.34 J.-M. Reau, J. Portier: In: *Solid Electrolytes*, ed. by P. Hagenmuller, W. Van Gool (Academic, New York 1978) pp. 313–333
- 11.35 H. A. Harwig, A. G. Gerards: *J. Solid State Chem.* **26**, 265–274 (1978)
- 11.36 T. Takahashi, H. Iwahara: *Mater. Res. Bull.* **1**(3), 1447–1453 (1978)
- 11.37 K. R. Kendall, C. Navas, J. K. Thomas, H.-C. zur Loye: *Chem. Mater.* **8**, 642–649 (1996)
- 11.38 J. C. Boivin, G. Mairesse: *Chem. Mater.* **1**(0), 2870–2888 (1998)
- 11.39 F. Abraham, J. C. Boivin, G. Mairesse, G. Nowogrocki: *Solid State Ionics* **4**(0–1), 934–937 (1990)
- 11.40 H. L. Tuller, D. P. Button, D. R. Uhlmann: *J. Non-Cryst. Solids* **42**, 297–306 (1980)
- 11.41 H. L. Tuller, M. W. Barsoum: *J. Non-Cryst. Solids* **73**, 331–50 (1985)
- 11.42 F. A. Fusco, H. L. Tuller: In: *Superionic Solids and Solid Electrolytes: Recent Trends*, ed. by A. L. Laskar, S. Chandra (Academic, New York 1989) pp. 43–110
- 11.43 D. Ravaine, J. L. Souquet: *Phys. Chem. Glasses* **18**, 27–31 (1977)
- 11.44 D. P. Button, R. P. Tandon, H. L. Tuller, D. R. Uhlmann: *J. Non-Cryst. Solids* **42**, 297–306 (1980)
- 11.45 D. P. Button, R. P. Tandon, H. L. Tuller, D. R. Uhlmann: *Solid State Ionics* **5**, 655–658 (1981)
- 11.46 F. A. Fusco, H. L. Tuller, D. P. Button: In: *Proc. Symp. Electro-Ceramics and Solid State Ionics*, ed. by H. L. Tuller, D. M. Smyth (Electrochemical Society, Pennington 1988) pp. 167–178
- 11.47 D. P. Button, P. K. Moon, H. L. Tuller, D. R. Uhlmann: *Glastech. Ber.* **56K**, 856–861 (1983)
- 11.48 M. A. Ratner, P. Johansson, D. F. Shriver: *MRS Bull.* **25**, 31–36 (2000)
- 11.49 B. Scrosati, C. A. Vincent: *MRS Bull.* **25**, 28–30 (2000)
- 11.50 D. Y. Wang, D. S. Park, J. Griffiths, A. S. Nowick: *Solid State Ionics* **2**, 95–105 (1981)
- 11.51 T. Ishihara, H. Matsuda, Y. Takita: *J. Am. Chem. Soc.* **11**(6), 3801–3803 (1994)
- 11.52 H. Iwahara, T. Esaka, H. Uchida, H. Maeda: *Solid State Ionics* **3**(4), 359–363 (1981)
- 11.53 A. S. Nowick, Y. Du: *Solid State Ionics* **7**(7), 137–146 (1995)
- 11.54 M. Cherry, M. S. Islam, J. D. Gale, C. R. A. Catlow: *J. Phys. Chem.* **9**(9), 14614–14618 (1995)
- 11.55 D. Y. Wang, D. S. Park, J. Griffiths, A. S. Nowick: *Solid State Ionics* **2**, 95–105 (1981)
- 11.56 R. Gerhart-Anderson, A. S. Nowick: *Solid State Ionics* **5**, 547–550 (1981)

- 11.57 H. L. Tuller, A. S. Nowick: *J. Electrochem Soc.* **126**, 209–217 (1979)
- 11.58 C. C. Liang: *J. Electrochem Soc.* **120**, 1289 (1973)
- 11.59 H. Sata, K. Eberman, K. Eberl, J. Maier: *Nature* **408**, 946–48 (2000)
- 11.60 J. Maier: *Prog. Solid State Chem.* **23**, 171–263 (1995)
- 11.61 Y.-M. Chiang, E. B. Lavik, I. Kosacki, H. L. Tuller, J. Y. Ying: *J. Electroceram.* **1**, 7–14 (1997)
- 11.62 H. Seh, *Langasite Bulk Acoustic Wave Resonant Sensor for High Temperature Applications*, PhD thesis, Dept. Materials Sc. & Eng. MIT, February, 2005.
- 11.63 L. Heyne: In: *Solid Electrolytes*, ed. by S. Geller (Springer, Berlin, Heidelberg 1977) p.169
- 11.64 B. M. Kulwicki, S. J. Lukasiewicz, S. Subramanyam, A. Amin, H. L. Tuller: In: *Ceramic Materials for Electronics*, 3rd edn., ed. by R. C. Buchanan (Marcel Dekker, New York 2004) pp. 377–430
- 11.65 N. Q. Minh, T. Takahashi: *Science and Technology of Ceramic Fuel Cells* (Elsevier, Amsterdam 1995)
- 11.66 H. L. Tuller: In: *Oxygen Ion and Mixed Conductors and their Technological Applications*, ed. by H. L. Tuller, J. Schoonman, I. Riess (Kluwer, Dordrecht 2000) pp. 245–270
- 11.67 C. Julien, G.-A. Nazri: *Solid State Batteries: Materials Design and Optimization* (Kluwer, Boston 1994)
- 11.68 C. G. Granqvist: In: *The CRC Handbook of Solid State Electrochemistry*, ed. by P. J. Gellings, H. J. M. Bouwmeester (CRC, Boca Raton 1997) pp. 587–615
- 11.69 P. T. Moseley, B. C. Tofield (eds.): *Solid State Gas Sensors* (Adam Hilger, Bristol 1987)
- 11.70 W. Göpel, T. A. Jones, M. Kleitz, I. Lundström, T. Seiyama (eds.): *Sensors: A Comprehensive Survey, Chemical and Biochemical Sensors*, Vol. 2nd and 3rd (VCH, New York 1991)
- 11.71 A. D. Brailsford, M. Yussouff, E. M. Logothetis: *Sensor. Actuat. B* **44**, 321–326 (1997)
- 11.72 J. Maier, M. Holzinger, W. Sitte: *Solid State Ionics* **74**, 5–9 (1994)
- 11.73 N. Yamazoe, N. Miura: *J. Electroceram.* **2**, 243–255 (1998)
- 11.74 T. Takeuchi: *Sensor. Actuat. B* **14**, 109–124 (1988)
- 11.75 W. Göpel, G. Reinhardt, M. Rosch: *Solid State Ionics* **136–137**, 519–531 (2000)
- 11.76 S. C. Singhal: *MRS Bull.* **25**, 16–21 (2000)
- 11.77 H. J. M. Bouwmeester, A. J. Burggraaf: In: *The CRC Handbook of Solid State Electrochemistry*, ed. by P. J. Gellings, H. J. M. Bouwmeester (CRC, Boca Raton 1997) p. 481
- 11.78 I. Riess, M. Godickemeier, L. J. Gauckler: *Solid State Ionics* **90**, 91–104 (1996)
- 11.79 S. B. Adler, J. A. Lane, B. C. H. Steele: *J. Electrochem. Soc.* **143**, 3554–3564 (1996)
- 11.80 S. A. Kramer, M. A. Spears, H. L. Tuller: Novel Compatible Solid Electrolyte–Electrode System Suitable for Solid State Electrochemical Cells, U. S. Patent No. 5,540,346 (1995)
- 11.81 N. J. Long, F. Lecarpentier, H. L. Tuller: *J. Electroceram.* **3:4**, 399–407 (1999)
- 11.82 F. Lecarpentier, H. L. Tuller, N. Long: *J. Electroceram.* **5**, 225–230 (2000)
- 11.83 B. Ma, U. Balachandran: *J. Electroceram.* **2**, 135–142 (1998)
- 11.84 T. J. Mazanec: *Solid State Ionics* **70/71**, 11–19 (1994)
- 11.85 K. Huang, M. Schroeder, J. B. Goodenough: *J. Electrochem. Solid State Lett.* **2**, 375–378 (1999)
- 11.86 J. E. ten Elshof, N. Q. Nguyen, M. W. den Otter, H. J. M. Bouwmeester: *J. Electrochem. Soc.* **144**, 4361–4366 (1997)
- 11.87 H. Takamura, K. Okumura, Y. Koshino, A. Kamegawa, M. Okada: *J. Electroceram.* **13**, 613 (2004)
- 11.88 M. Lazzari, B. Scrosati: *J. Electrochem. Soc.* **127**, 773–774 (1980)
- 11.89 J. R. Dahn, A. K. Sleight, H. Shi, B. M. Way, W. J. Weydanz, J. N. Reimers, Q. Zhong, U. von Sacken: In: *Lithium Batteries*, ed. by G. Pistoia (Elsevier, Amsterdam 1994) pp. 1–47
- 11.90 J.-M. Tarascon, W. R. McKinnon, F. Coowar, T. N. Bowmer, G. Amatucci, D. Guyomard: *J. Electrochem. Soc.* **141**, 1421–1431 (1994)
- 11.91 C. Julien, G.-A. Nazri: *Solid State Batteries: Materials Design and Optimization* (Kluwer, Boston 1994)
- 11.92 F. M. Gray: *Solid Polymer Electrolytes: Fundamentals and Technological Applications* (VCH, New York 1991)
- 11.93 C. M. Lampert: *Solar Energy Mat.* **11**, 1–27 (1984)
- 11.94 M. Green: *Ionics* **5**, 161–170 (2000)

Growth a

Part B

Part B Growth and Characterization

- | | |
|---|---|
| <p>12 Bulk Crystal Growth – Methods and Materials
Peter Capper, Southampton, UK</p> <p>13 Single-Crystal Silicon: Growth and Properties
Fumio Shimura, Fukuroi, Japan</p> <p>14 Epitaxial Crystal Growth: Methods and Materials
Peter Capper, Southampton, UK
Stuart Irvine, Gwynedd, UK
Tim Joyce, Liverpool, UK</p> <p>15 Narrow-Bandgap II–VI Semiconductors: Growth
Peter Capper, Southampton, UK</p> | <p>16 Wide-Bandgap II–VI Semiconductors: Growth and Properties
Jifeng Wang, Sendai, Japan
Minoru Isshiki, Sendai, Japan</p> <p>17 Structural Characterization
Paul D. Brown, Nottingham, UK</p> <p>18 Surface Chemical Analysis
David Sykes, Loughborough, UK</p> <p>19 Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications
Safa Kasap, Saskatoon, Canada
Dan Tonchev, Saskatoon, Canada</p> <p>20 Electrical Characterization of Semiconductor Materials and Devices
M. Jamal Deen, Hamilton, Canada
Fabien Pascal, Montpellier, France</p> |
|---|---|

12. Bulk Crystal Growth – Methods and Materials

This chapter covers the field of bulk single crystals of materials used in electronics and optoelectronics. These crystals are used in both active and passive modes (to produce devices directly in/on bulk-grown slices of material, or as substrates in epitaxial growth, respectively). Single-crystal material usually provides superior properties to polycrystalline or amorphous equivalents. The various bulk growth techniques are outlined, together with specific critical features, and examples are given of the types of materials (and their current typical sizes) grown by these techniques. Materials covered range from Group IV (Si, Ge, SiGe, diamond, SiC), Group III–V (such as GaAs, InP, nitrides) Group II–IV (including CdTe, ZnSe, [MCT](#)) through to a wide range of oxide/halide/phosphate/borate materials. This chapter is to be treated as a snapshot only; the interested reader is referred to the remainder of the chapters in this Handbook for more specific growth and characterization details on the various materials outlined in this chapter. This chapter also does not cover the more fundamental aspects of the growth of the particular materials covered; for these, the reader is again referred to relevant chapters within the Handbook, or to other sources of information in the general literature.

Despite the widespread progress in several epitaxial growth techniques for producing electronic and optoelectronic device-quality material, various bulk growth methods are still used to produce tens of thousands of tons of such materials each year. These crystals are used in both active and passive modes; in other words, to produce devices directly in/on bulk-grown slices of material, or used as substrates in epitaxial growth, respectively.

This chapter covers the field of bulk single crystals of materials used in electronics and optoelectronics. Single-crystal material usually provides superior properties to polycrystalline or amorphous equivalents. The various bulk growth techniques are outlined, together with specific critical features, and examples are given

12.1	History	232
12.2	Techniques	233
12.2.1	Verneuil	233
12.2.2	Czochralski	233
12.2.3	Kyropoulos	234
12.2.4	Stepanov	235
12.2.5	Edge-Defined Film Growth	235
12.2.6	Bridgman	236
12.2.7	Vertical Gradient Freeze	237
12.2.8	Float Zone	237
12.2.9	Travelling Heater Method (THM) ..	237
12.2.10	Low-Temperature Solution Growth	238
12.2.11	High-Temperature Solution Growth (Flux)	238
12.2.12	Hydrothermal	239
12.2.13	Growth from the Vapor	240
12.3	Materials Grown	240
12.3.1	Group IV	240
12.3.2	Groups III–V	242
12.3.3	Groups II–VI	245
12.3.4	Oxides/Halides/Phosphates/ Borates	249
12.4	Conclusions	251
	References	251

of the types of materials grown by these techniques, as well as their current typical sizes. We cover materials ranging from Group IV (Si, Ge, SiGe, diamond, SiC), Group III–V (GaAs, InP, nitrides, among others) and Group II–IV (such as CdTe, ZnSe, [MCT](#)) materials through to various oxides/halides/phosphates/borates. However, this chapter should only be treated as a brief foray into the field. The reader interested in more detail is referred to the chapters following this one (or to other sources of information in the general literature) for more on the growth and characterization of various materials outlined in this chapter. Many of the crystals grown at small sizes, mainly for R&D purposes, particularly in universities, are not discussed; neither, in general, are the organic materials that are being studied in this area.

12.1 History

Several very useful studies on the history of crystal growth in general can be found in the literature [12.1–7]. Many significant contributions were made to the fundamentals of crystal growth during the eighteenth and nineteenth centuries, including the development of thermodynamics, undercooling and supersaturation [12.6, 7]. In terms of crystal growth techniques, it is accepted that the first method applied to produce usable crystals on a large scale was that of flame fusion by Verneuil [12.8].

Before World War II, synthetic crystals (other than ruby) were mainly used in scientific instruments. However, between 1900 and 1940 there were enormous advances in both the theoretical area of this field and in producing samples for scientific study. The diffusion boundary layer was applied by Nernst [12.9], while ideas on the growth of perfect crystals were proposed by Volmer [12.10], Kossel [12.11] and Stranski [12.12].

Table 12.1 Estimated worldwide annual production rates of crystals (as at 1986, [12.1]). Note that some materials are mainly used for non-(electronic/optoelectronic) purposes; for example cubic zirconia and much of the ruby are used in jewelry and watches (items in brackets are small in volume but high in value)

Crystal	Rate (t/yr)	Growth methods
Silicon	4000	Czochralski, float-zone, (VPE)
Metals	4000	Bridgman, Strain anneal
Quartz	800	Hydrothermal
III–V compounds	600	Czochralski, Bridgman (VPE, LPE)
Alkali halides	500	Bridgman, Kyropoulos
Ruby	500	Verneuil
Germanium	400	Czochralski, Bridgman
Garnets	200	Czochralski
Lithium niobate	100	Czochralski
Phosphates	50	Low-temperature solution
Lithium tantalate	20	Czochralski
Cubic zirconia	15	Skull melting
TGS	10	Low-temperature solution
Diamond	10	High-temperature solution
II–VI compounds	5	Vapor, Bridgman

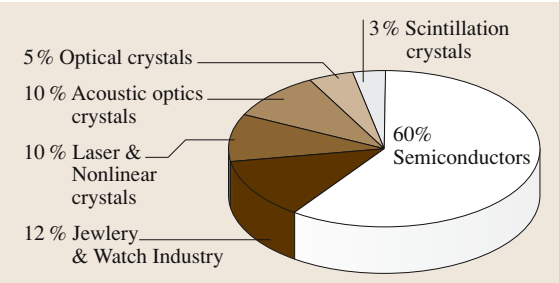


Fig. 12.1 Estimated share of the world production of 20 000 t of bulk crystals (1999). After [12.2], copyright Elsevier Science (used with permission)

Many of the growth techniques now used were also initially developed during this period. The flame fusion of Verneuil [12.8] was followed by hydrothermal growth [12.13, 14], crystal pulling [12.15], Kyropoulos [12.16], Bridgman growth [12.17–20], and vertical gradient freeze [12.21].

During World War II synthetic crystals were used as piezo-electric transducers (in sonar) as resonant devices (in radar) or as infra-red detectors. However, it was the invention of the transistor (announced in 1948) that heralded the modern era of crystal growth for practical purposes rather than for pure scientific interest. Teal and Little developed crystal pulling of Ge in 1950 [12.22] and Dash [12.23] improved this with the ‘necking’ technique to produce dislocation-free material. A key issue at this stage was measuring levels of impurities. The semiconductors used in solid-state devices such as transistors work because they have selected dopants in them. Techniques were developed at this time to improve the purity of materials, such as zone refining [12.24] and float-zone refining [12.25, 26]. Theoretical aspects developed in this period include the role of screw dislocations in growth [12.27] and the generalized theory of Burton et al. [12.28]. Reducing melt inclusions in crystals was discussed by Ivantsov [12.29, 30] in terms of diffusional undercooling, and by Tiller et al. [12.31] in terms of constitutional supercooling. Similar work was carried out in solution growth by Carlson [12.32], who studied flow effects across crystal faces, and Scheel and Elwell [12.33], who derived the maximum stable growth rate and optimized supersaturation to produce inclusion-free crystals. Segregation effects, which are related to mass and heat transfer, were studied by Burton et al. [12.34] for melt growth and by van Erk [12.35] for solution growth,

while the experimental conditions needed to produce striation-free material were established by Rytz and Scheel [12.36]. Forced convection in diffusion-limited growth was recognized as being beneficial for open systems with stirrers [12.37–40], while stirring in sealed containers was accomplished using the accelerated crucible rotation technique [12.41].

In 1986, Brice [12.1] estimated the annual production rates of crystals, and Table 12.1 is a reproduction of that data. Semiconductor materials clearly dominated at that stage, in particular silicon and III–V compounds. Brice also tabulated the uses of the various crystals. Later, Scheel [12.6] gave estimates of

5000 t in 1979, 11 000 t in 1986 (from [12.1]) and approximately 20 000 t in 1999, see Fig. 12.1. By 1999 the balance had shifted somewhat from the earlier estimates, but semiconductors continued to dominate, at $\approx 60\%$. There were roughly equal percentages of scintillator, optical and acousto-optical crystals, at around 10–12%. The remainder were made up of laser and non-linear optical crystals and crystals for jewelry and the watch industries.

A recent book edited by Capper [12.42] gives a comprehensive update on the bulk growth of many of the materials used in the electronic, optical and optoelectronic fields.

12.2 Techniques

12.2.1 Verneuil

This is the fastest growth method and was the first found to be capable of controlling nucleation and thus producing large crystals of high melting point oxide crystals, such as sapphire and ruby. Currently, a large number of high melting point materials have been grown by this technique, including ZrO_2 (2700 °C), SrO (2420 °C) and Y_2O_3 (2420 °C), but the largest use is still Al_2O_3 (often doped to produce ruby, sapphire, and so on). Figure 12.2 shows a schematic of the equipment used. Many different heat sources have been used, such as solar furnaces, glow discharges, plasma torches, arc images and radio-frequency heating, but the original gas flame technique is still the most popular. An oxy-hydrogen flame heats the seed crystal, which is placed on a ceramic pedestal. Powder from a hopper is shaken through the flame and melts, forming a melt surface on the seed. During growth, the seed is lowered, controlling the linear growth rate, while the volume growth rate is governed by the powder feed rate. The balance of these two rates controls the crystal diameter and the crystal is normally rotated slowly. Ruby crystals up to 200 mm in diameter can be grown. Drawbacks of the technique include high dislocation densities and concentration variations.

12.2.2 Czochralski

Czochralski is a fast growth method widely used for both semiconductors and oxide/fluoride materials intended for optical applications. It usually also produces the most homogeneous crystals and those with the fewest flaws. However, it is only really applicable to those materials

that melt congruently or nearly congruently, so that the solid and the melt compositions are the same at equilibrium. A crucible material is normally needed that is compatible with the melt, but the crucible material most commonly used for silicon – silica – dissolves slowly in the melt and this raises certain process issues about

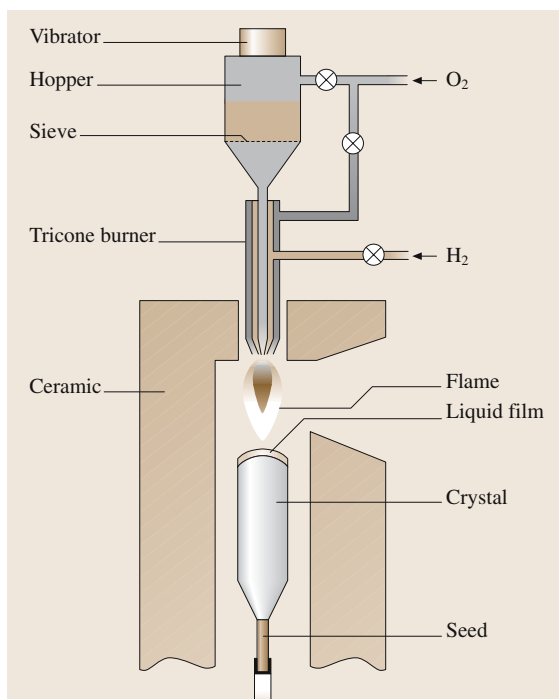


Fig. 12.2 Schematic of Verneuil growth equipment. (After [12.1])

the growth of this material. Capital costs for Czochralski pulling are higher than most other techniques, but it is used when the greatest perfection is required.

The largest use is clearly for silicon, but ≈ 100 materials are grown commercially using this method, while many more are at the research stage. As the diameter of the crystal increases, for example from 200 mm to 300 mm for silicon, controlling dislocations becomes more difficult due to increased radial temperature gradients. Figure 12.3 shows a schematic of Czochralski growth equipment. The basic method is relatively simple: the solid charge is placed in the crucible and heated to a temperature several degrees above the melting point. The seed crystal, rotating slowly, is lowered to contact the melt and the seed then slowly melts. After a short delay (a few minutes), pulling is commenced and new material (with slightly reduced diameter) begins to grow. A long narrow 'neck' region is grown to reduce dislocations, and then the melt temperature is reduced to increase the diameter. When the crystal attains full diameter, growth is maintained until the desired length of crystal has grown, and growth is terminated by sharply increasing the pull rate or increasing the melt temperature so that the diameter reduces to zero.

Constant diameter is maintained by adjusting the power input to the melt. This is done automatically, ei-

ther by directly monitoring the diameter optically (by observing the bright ring around the crystal periphery), or by indirectly measuring the diameter via weighing methods. Crucibles are normally round-based for semi-conductors but should be relatively flat for oxide/halide growth (for ionic materials). Often the depth of the initial melt is the same as or slightly less than the diameter of the crystal.

Hurle [12.43] discusses the effects, in detail, of convection, flows in melts, heat transport, mass transport, solute segregation, use of magnetic fields, systems dynamics and automatic diameter control, morphological stability and defect control. In general, as melt sizes have increased natural convective flows have also increased, leading to turbulence, which causes growth rate fluctuations leading to dopant concentration variations (so-called "striations").

For the growth of III–V compounds, liquid-encapsulated Czochralski (LEC) was pioneered by *Mullin* and coworkers [12.44] for GaAs and GaP. Suppressing the volatility of As and P was crucial to the successful growth of these types of compounds. This technique involves the use of an inert layer of a transparent liquid, usually B_2O_3 , which floats on the melt surface and acts as a liquid seal. Most importantly, the encapsulant should wet both the crucible and crystal so that a thin film adheres to the crystal as growth proceeds. The latter prevents dissociation of the hot crystal. PBN crucibles are often used these days, and either resistance or RF heating is employed. Semi-insulating GaAs is produced in a high-pressure puller (100–200 atm). The main advantage here is that elemental Ga and As can be used as the starting materials. Diameter is often controlled via crystal-weight measurements, rather than by meniscus observation. Pull rates are typically < 1 cm/h. Axial or transverse magnetic fields can again be used to control melt turbulence and dopant segregation.

Precious metal (Pt, Ir) crucibles are used for oxide crystal growth, as high temperatures are used. To prevent the reduction of oxide melts, a partial pressure of oxygen is needed. Normally, the crucible is not rotated or translated and the growth rate is greater than the pulling rate as the melt height decreases. Stable growth rates are obtained for crystal diameters of less than half the crucible diameter. Growth rates are a few mm/h at best, so growth times are normally > 1 week. Large temperature gradients can lead to melt turbulence, which in turn produces a banded structure of solute concentration and varying stoichiometry in the crystals. Controlling convection in these systems is therefore of critical importance.

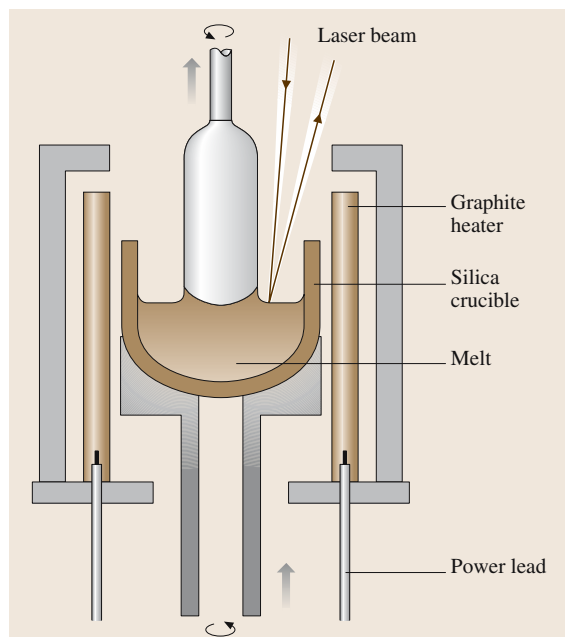


Fig. 12.3 Schematic of Czochralski growth equipment. (After [12.1])

12.2.3 Kyropoulos

This simple technique is used where a large diameter is more important than the length, as in the cases of windows, prisms, lenses and other optical components, and for scintillator materials. The set-up and method of growth are similar to those of Czochralski (Fig. 12.4), but after the seed is brought into contact with the melt it is not raised much during growth. As in Czochralski, a short necked region is still grown but then the seed removal is stopped and growth proceeds by reducing the input power to the melt. The resulting crystals normally have diameters of $\approx 80 - 90\%$ of the crucible diameter. Although the control systems used for this process are relatively simple, the temperature distribution over the crucible is critical. For alkali halides the crystal density is greater than the melt density, so the melt level decreases with growth, and the desired temperature distribution is one of increasing temperature as the base of the crucible is approached. Average linear growth rates are a few mm/h, corresponding to cooling rates of $< 1^\circ\text{C/h}$. The only other process parameter of concern is the seed rotation rate, which is normally low or zero.

While the method appears to be very attractive economically, there are technical deficiencies. The isothermal surfaces are curved, resulting in high dislocation densities, and the growth interface is composed of different crystal faces, with consequent inhomogeneities in impurities and vacancy concentrations. High levels of impurities result from the majority of the melt (which is not stirred well) being consumed. Despite these drawbacks, many tons of alkali halides are grown for optical applications each year by this technique.

12.2.4 Stepanov

In this technique, a crystal is pulled from a crucible containing a crystal-shaped aperture (Fig. 12.5). Crystals can be pulled vertically upwards, downwards, or even horizontally. Growth rates are below those of the normal Czochralski technique, but dislocation densities can be reasonably low.

12.2.5 Edge-Defined Film Growth

A die with a central capillary is placed on the surface of the melt (Fig. 12.6). Surface tension forces cause the melt to wet the die and be drawn up the capillary. A pointed seed crystal is lowered to contact this melt and then pulled upwards. The melt is cooled slightly to increase the crystal diameter until it reaches the size of

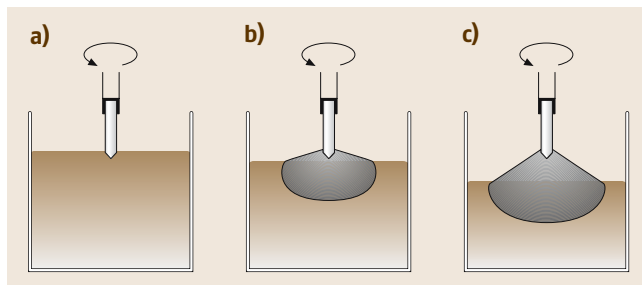


Fig. 12.4a–c Schematic of Kyropoulos growth equipment. (a) The seed crystal contacts the melt, a small amount melts and then cooling is commenced to produce (b) and (c). (After [12.1])

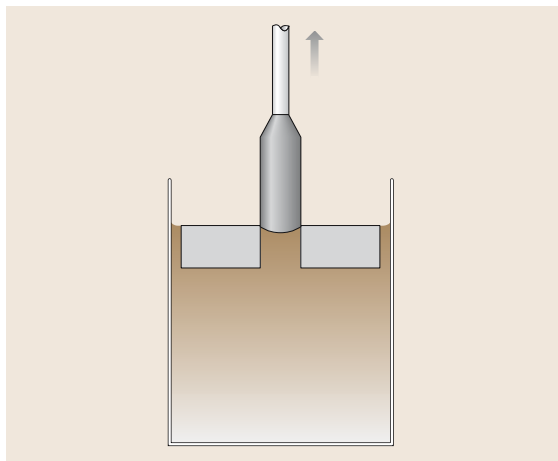


Fig. 12.5 Schematic of Stepanov growth equipment, in which a crystal is pulled through an aperture that defines its shape. (After [12.1])

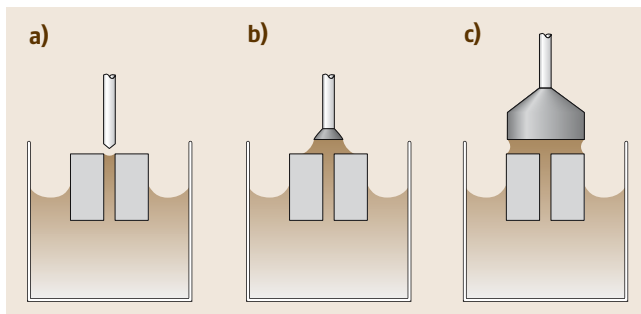


Fig. 12.6a–c Schematic of edge-defined film growth equipment. (a) The melt wets the die and is drawn up the capillary; (b) the seed is contacted to the melt and pulled up and cooling starts to increase the crystal size; (c) the crystal reaches the size of the die. (After [12.1])

the die. Die can be designed to produce various shapes of crystals, including tubes and sheets. Rapid growth is possible, but crystal quality normally suffers. One product is alumina tape (1 mm thick by several centimeters wide) used as a substrate for the production of high-frequency circuits. In this technique, however, purity can be limited.

12.2.6 Bridgman

In essence, this is a method of producing a crystal from a melt by progressively freezing it from one end to the other (Fig. 12.7). Crystals can be obtained with good dimensional control and the method uses relatively simple technology requiring little supervision. However, as the diameter increases, controlling heat flow becomes progressively more difficult. A wide range of materials have been produced by this technique, including sapphire at a melting point of $\approx 2370^\circ\text{C}$. One major requirement is that neither the melt nor its vapor must attack the crucible material significantly. Dislocation densities can also be limited to $> 10^4 \text{ cm}^{-2}$, and many materials contain low-angle grain boundaries.

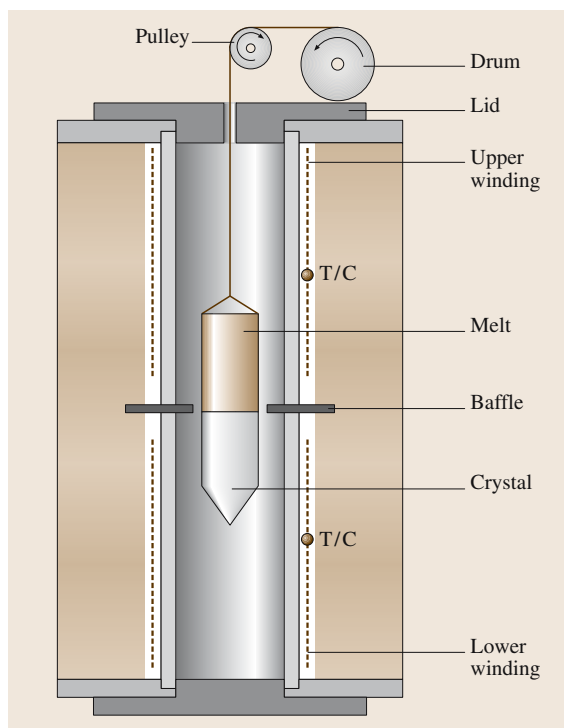


Fig. 12.7 Schematic of Bridgman growth equipment. (After [12.1])

Growth rates can be in the range 0.1–30 mm/h. Either the crucible or the furnace can be moved to achieve movement of the freezing isotherm, and both vertical and horizontal orientations are used. If vibration is an issue in a particular growth system, then it is usually preferable to move the furnace rather than the crystal. Crucible materials include silica (for covalently bonded crystals), graphite and some metals (such as Mo for sapphire growth). The crucibles traditionally have tapered tips in order to try to restrict nucleation to one crystal, although seed crystals can also be used. If spurious nucleation occurs it usually forms at the crucible walls, but these can be suppressed by making the growth face concave into the melt. Baffles are often used to separate upper and lower parts of the furnace to ensure thermal isolation. A low radial temperature gradient is needed to reduce dislocation densities. Calculating the temperature distribution is possible [12.45]. The most obvious requirements are a large temperature gradient at the growth interface and low temperature gradients in the radial direction.

Impurity distributions in crystals are, in general, governed by the type of mixing in the melt. In stirred melts – so-called normal freezing – the relevant equation is

$$C_s = k_{\text{eff}} C_0 (1 - x) k_{\text{eff}} - 1,$$

where C_s = concentration in the solid, C_0 = initial concentration, k_{eff} = effective distribution coefficient, and x is the fraction solidified. In unstirred melts, the interface segregation coefficient k^* governs

$$C_s = k^* C_0 (1 - [(1 - k^*)/k] \times \{1 - \exp[-(1 - k^*)k^* f_z / D_L]\}),$$

where f is the growth rate, z the axial distance, and D_L is the diffusion coefficient. For a molten zone of length L , the equation is

$$C_s = C_0 [1 + (k_{\text{eff}} - 1) \exp(-k_{\text{eff}} x)].$$

For materials with a volatile component there are several possible refinements of the basic process, such as sealed ampoules, overpressure and liquid encapsulation (for example using B_2O_3). In sealed-ampoule growth there is normally a lack of control over stirring. The accelerated crucible rotation technique (ACRT) can be used for vertical systems, and this author has developed this refinement for the growth of cadmium mercury telluride up to 20 mm in diameter [12.46]. Faster stable growth rates are possible with ACRT [12.46], as are larger diameter crystals, improved uniformity and better crystallinity.

It is worth noting that the majority of melt-grown crystals are produced by the Bridgman and Czochralski techniques (and their variants). In general, if no suitable crucible can be found, then other methods are required, such as the Verneuil method.

12.2.7 Vertical Gradient Freeze

The vertical gradient freeze (VGF) technique involves the progressive freezing of the lower end of a melt upwards. This freezing process can be controlled by moving the furnace past the melt or, preferable, by moving the temperature gradient in a furnace with several independently controlled zones. Low-temperature gradients are normally obtained, leading to reduced dislocation densities, and the crystal is of a defined shape and size. Difficulties include furnace design, the choice of boat material, and the issue of seeding.

12.2.8 Float Zone

In this technique a molten zone is maintained between two solid rods (Fig. 12.8). By moving the zone relative to the rods, one of them grows, and a single crystal can be grown if a seed is used. Silicon is the

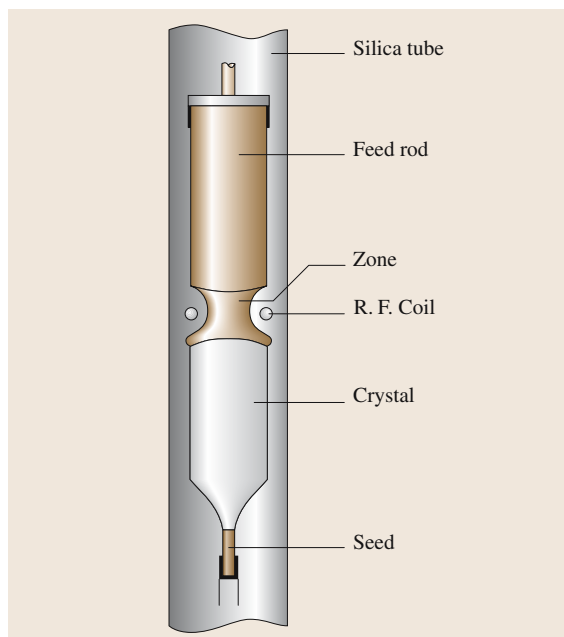


Fig. 12.8 Schematic of float-zone growth equipment. (After [12.1])

only material grown on a large scale by this technique. The only other use for it is in the small-scale growth of very pure crystals, as no crucible contact is involved. For silicon, RF heating is used, with frequencies of 2–3 MHz for diameters > 70 mm. There is a steep temperature gradient that induces flows in the molten zone, and if both the seed and feed rods are rotated then the shape of the solid/liquid interface can be controlled. The molten zone normally moves upwards.

12.2.9 Travelling Heater Method (THM)

This technique was developed for II–VI alloy growth (Te-based) by Triboulet [12.47]. In the technique, a molten zone is made to migrate through homogeneous solid source material. This is normally accomplished by slowly moving the ampoule relative to the heater (Fig. 12.9). The key requirement here is to obtain the appropriate temperature profile, also shown in Fig. 12.9. Matter transport is by convection and diffusion across the solvent zone under the influence of the temperature gradient resulting from the movement. For alloy growth, a steady state can be reached where the solvent dissolves a solid of composition C_0 at the upper growth interface and deposits, at near-equilibrium conditions, a material of the same composition at the lower growth interface. Growth occurs at a constant temperature below the solidus temperature and hence shows all the advantages of low-temperature growth. When tellurium was used as the molten zone there

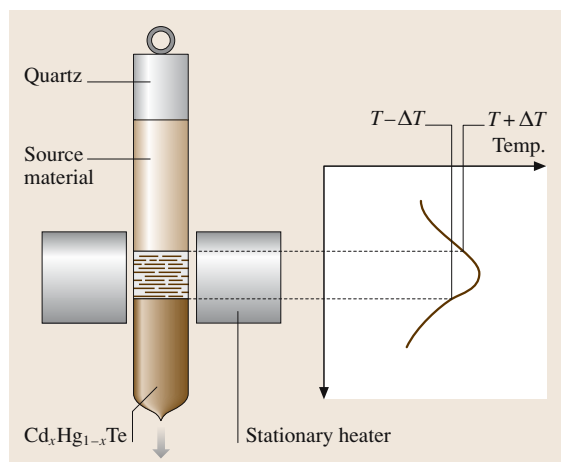


Fig. 12.9 Schematic of travelling heater method growth equipment. (After [12.1])

was a marked purification effect. Seeding was also found to be possible with the technique, as was growth to a larger diameter than achieved with the other common melt techniques used for these compounds.

Problems with THM include the availability of suitable feed material with the required composition and dimensions, although several different routes have been used to overcome this problem. Natural convection is the dominant mechanism of material transport, which led to the addition of ACRT to the basic THM method for some of the ternary alloys.

Graded composition alloys were also made by producing two bevelled cylinders of binary compounds. These were then used to assess the effects of composition on various electrical and optical properties in the given system. Other modifications to the basic process included the 'cold THM' process, in which the relevant metallic elements were used as the source material. This produced a process that accomplished synthesis, growth and purification at low temperature all in a single run. To avoid the problems resulting from solvent excess in the crystal, 'sublimation THM' was developed, in which an empty space of the same dimension as the molten zone is used. This method was successfully applied to the growth of ZnSe. Repeated runs on the same material were used to improve the purification effect. A drawback of this is repeated handling, which in turn was solved by using the 'multipass THM' technique, which can be thought of as a variety of the classical zone-melting method.

12.2.10 Low-Temperature Solution Growth

Most of the crystals grown by this technique are water soluble. This limits their use to applications in which moisture can be excluded. Growth rates are low (0.1–10 mm/d), as the growth faces are unstable. This is due to the concentration gradient near the growth face, in addition to which the crystal is normally totally immersed in the solution so that latent heat evolved increases the adverse supersaturation gradient (Fig. 12.10).

Water (both light and heavy) is used in $\approx 95\%$ of the cases, and must be highly pure. All equipment must be carefully cleaned and protected from dust and the solutions must be stirred vigorously while being prepared. Both slow cooling and solvent evaporation techniques are used. Large volumes of solution (> 100 l) are needed for very large crystals of, for example, KH_2PO_4 (KDP). These large volumes imply very long thermal time

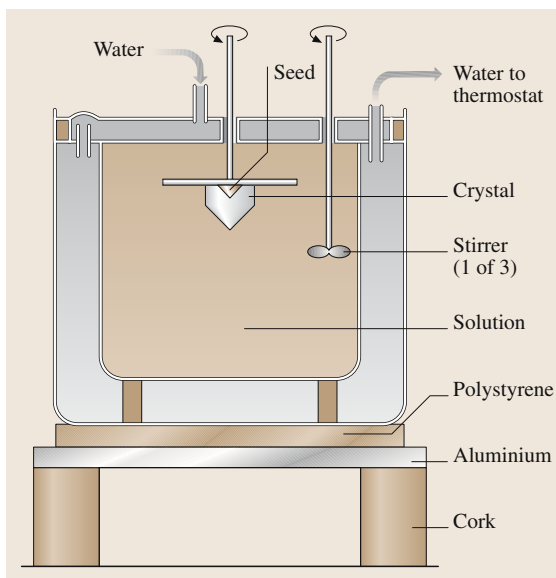


Fig. 12.10 Schematic of low-temperature solution growth equipment. (After [12.1])

constants, which can be increased by adding thermal insulation, up to a limit. Large time constants make it easy to obtain good temperature control. The growth equipment is normally held in temperature-controlled rooms. Although temperature stability is key, other parameters, such as fluctuations in stirring, can lead to large changes in growth rate too.

Solvent evaporation simplifies the temperature control system and makes it more reliable. However, it is difficult to ensure a constant acceptable rate of loss of solvent. This can be approached by using a water-cooled condensation region. One drawback of this technique is that concentrations of impurities increase as the growth proceeds.

12.2.11 High-Temperature Solution Growth (Flux)

This method is used for those materials that melt incongruently; the solvent (flux) reduces the freezing point below the relevant temperature to produce the desired phase of the compound. Both liquid metals (such as Ga, In, Sn) used in semiconductors and oxides/halides (such as PbO and PbF_2) used for ionic materials are employed as solvents. Often, an excess of one of the components will be used (Ga for GaAs and GaP for instance) or a common ionic material (such as K_2CO_3 for $\text{KTa}_x\text{Nb}_{1-x}\text{O}_3$). Alternatively, the solvents con-

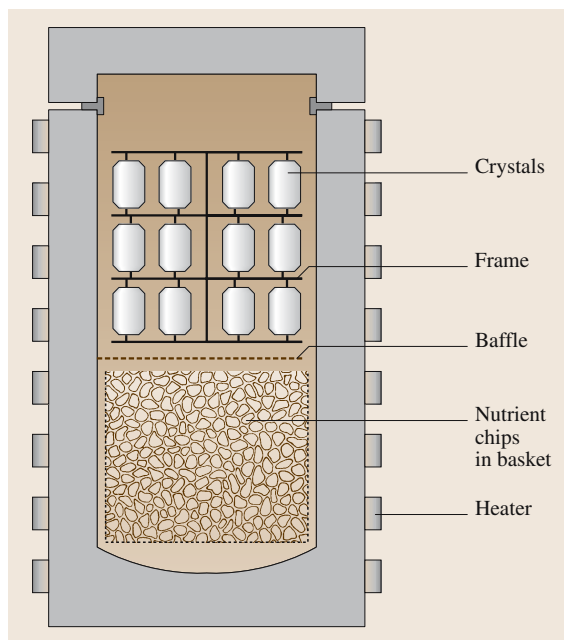


Fig. 12.11 Schematic of hydrothermal growth equipment. (After [12.1])

tain large atoms, such as Pb and Bi, which are too large to enter the lattice of the desired crystal. Mixed solvents, such as alkali metal ions, break the chains that exist in B_2O_3 , SiO_2 and Bi_2O_3 , lowering viscosities and surface tensions. The book by *Elwell and Scheel* [12.5] contains much useful information about this topic.

Slow cooling of high-temperature solutions was used between 1950 and 1970 to produce hundreds of different materials. Small crystals were normally obtained but these were sufficient to obtain useful measurements of magnetic, optical and dielectric properties. Later, seeded growth and stirring using *ACRT* led to much larger crystals. A Czochralski-type pulling technique can also be used in high-temperature solution growth to produce larger crystals.

Diamond is produced at high temperature ($> 1500\text{ K}$) and high pressure ($> 50\text{ kbar}$). Solvents used include Pt, Pd, Mn, Cr and Ta, but better results were obtained with Ni, Co and Fe and their alloys.

12.2.12 Hydrothermal

This is growth from aqueous solution at high temperature and pressure. Most materials grown have low solubilities in pure water, so other materials (called

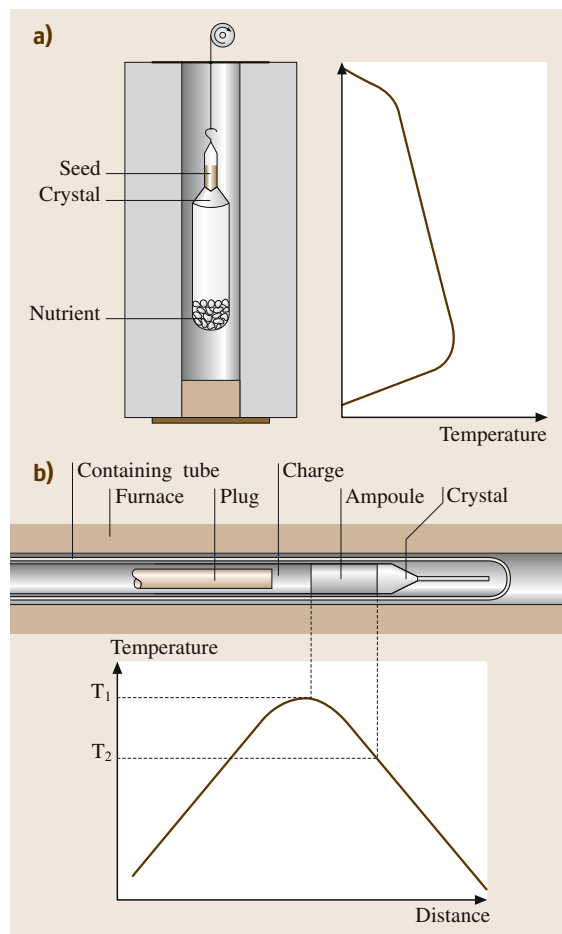


Fig. 12.12a,b Schematic of vapor growth equipment. (a) Seeded growth; (b) Piper-Polich set-up. (After [12.1])

mineralizers, such as $NaOH$ and Na_2CO_3) are added to increase the solubility. Quartz is the only material grown on a significant scale by this method, although ZnO is also produced commercially by this technique.

Natural quartz crystals are used as seeds, as they have lower dislocation densities. Growth proceeds in the temperature gradient transfer mode. Nutrient is held in the lower part of the liquid in an autoclave. This region is held at between 5 and 50°C above the upper portion of the autoclave. Convection carries the hot supersaturated liquid to the cooler upper regions, where deposition occurs on the seed crystals (Fig. 12.11). These autoclaves can be operated at up to 400°C and 2000 atm pressure for approximately 10 years without degradation. Autoclaves can be up to 1 m in diameter and 2.5 m high

for low-pressure growth from Na_2CO_3 solution, while 40-cm-diameter and 8-m-high autoclaves are typical of growth from NaOH .

Resistive heating elements are strapped to the outside of the autoclave and connected to two control systems, allowing independent control of the temperatures in the solution and growth zones. *Brice* [12.1] presents a table that shows some typical hydrothermal growth conditions for a wide range of materials.

The essential problem in the growth of quartz is to produce useful devices with reasonable yields with OH concentration (measured by IR absorption) of < 100 ppm. Several parameters restrict the maximum growth rate to ≈ 0.5 mm/d.

12.3 Materials Grown

This section attempts to summarize the current position regarding the wide range of materials used in electronics and optoelectronics. Again the reader is referred to the following relevant chapters for more in-depth coverage of a particular material. The section is subdivided roughly into Group IV, Group III–V, Group II–VI, oxides, halides, and finally phosphates/borates.

12.3.1 Group IV

Silicon and Germanium

The growth of silicon still dominates the electronics industry in terms of size of activity. *Hurle* [12.43] describes in detail the growth of silicon by the Czochralski method. In essence a polysilicon charge is placed in a high-purity quartz crucible mounted in prebaked graphite supports. Chains or cables are used in the pulling system, rather than rods, to limit the height of modern pullers. Both the crystal and the crucible can be rotated, although the details of this aspect vary from manufacturer to manufacturer and are proprietary. As the growth proceeds, the crucible is raised to maintain the position of the melt surface in the heater.

Silicon has two major advantages from a crystal growth perspective: a high thermal conductivity (which permits the latent heat of solidification to be removed) and a high critical resolved shear stress (which allows high thermal gradients without dislocation generation). These two factors enable high growth rates, leading to economic benefits. Heat is provided by ‘picket fence’-type graphite heaters. Most commercial pullers are highly automated with full computer control. In most cases, diameter control is achieved via monitoring of the

12.2.13 Growth from the Vapor

Vapor growth has some applications, particularly in high melting point materials and in II–VI compounds such as ZnSe and ZnS (Fig. 12.12). High-temperature vapor transport, or sublimation, of SiC has progressed recently and is now able to produce 1–2-inch crystals. Growth proceeds by sublimation of a SiC source at 2000–2500 °C in near-vacuum conditions onto a seed held at ≈ 1000 °C lower than the source. 6H-polytype material oriented at $\langle 0001 \rangle$ is produced. Growth rates are ≈ 2 mm/h. Remaining problems include impurities ($\approx 10^{16} \text{ cm}^{-3}$), voids, high dislocation densities and mixed polymorphism.

‘bright ring’ around the meniscus of the growing crystal. Oxygen is a particular problem and is introduced via erosion of the silica crucible by the melt. Convection moves the oxygen within the melt towards the crystal, where it is incorporated. Slightly reduced pressure over the melt can help to remove oxygen from the melt as silicon monoxide. A similar situation arises with carbon monoxide loss from the melt. Problems with convection have increased as the diameters of silicon crystals have increased over the past several decades. This convection controls the concentrations and distributions of oxygen, carbon and other unwanted impurities.

Although current production of large-diameter material is at the 300 mm level, there have been recent reports of 400 mm-diameter material [12.48]. In this case it was necessary to grow a secondary neck below the Dash dislocation-reducing one, as the thin Dash neck could not hold the weight of the crystal. Crystals of up to 438 kg and 1100 mm in length were reported. A recent report [12.49] shows how codoping Si with B and Ge can obviate the need for this narrow neck region, while maintaining dislocation-free growth.

The highest grade of silicon (the highest purity) is probably still produced by the float-zone method, rather than the Czochralski technique [12.50, 51]. The higher purity (particularly regarding carbon and oxygen) and better microdefect control result in higher solar cell efficiencies in FZ material. Faster growth rates and heating/cooling times together with the absence of crucibles and hot-zone consumables also give FZ material an economic advantage over Czochralski material.

Table 12.2 summarizes the present position on sizes of various crystals, including silicon and germanium.

Table 12.2 Typical sizes (diameters) of a selection of bulk-grown crystals (in mm and/or weight in kg) currently in production and used in R&D

Crystal	Commercial	R&D
Silicon (Cz)	300 (250 kg)	400 (438 kg)
Silicon (FZ)	100	
Germanium	75	200–300
Silicon/Germanium		35 (120 mm long)
SiC	50	100
Diamond	10	10 × 10
GaAs	150 (50 kg)	200
InP	100	150
GaSb	50	75
InSb	75	100
InAs	50	
InGaAs		10
GaN (sublimation)	50	
MCT	20	40
CdZnTe	140 (10 kg)	140 (12 kg)
Bi ₂ Te ₃	12	28
ZnSe (vapor, LEC)	60	
ZnTe	80	
ZnO	50	
HgMnTe	30	
HgZnTe	30	
LiNbO ₃		40
PbZnNbTiO ₃		75
YAG	20	75
LiTaO ₃	100	125
Ruby	200	
BGO		130
La ₃ Ga ₅ SiO ₁₄		100
Quartz	15 – 30 × 150 – 250, (0.5–18 kg)	
CaF ₂	385 (100 kg)	
Alkali halides	500 (550 kg)	1000?
Phosphates	45 × 45 × 86	
TGS	40	

The author does not claim that this list is complete nor is it authoritative, it merely serves to show the range of crystals and sizes currently being grown by bulk growth techniques. The information is taken from a number of recently published conference proceedings [12.52–57] and other sources [12.42, 58–60].

Silicon Carbide

The growth of SiC for semiconductor devices, as opposed to abrasive applications, became successful when substrates could be reproducibly grown up to large sizes [12.61–63]. Currently, substrates up to 50 mm

in diameter are commercially available and 100 mm-diameter substrates have been reported. The current method is that of seeded sublimation growth. A water-cooled quartz reactor enclosure surrounds a graphite crucible, which is heated by means of RF. Crucible sizes are slightly larger than the required crystal size and SiC powder (particle size 20–200 μm) and/or sintered polysilicon is used as the source. A distance of 1–20 mm between the seed and the source is typically used. At the high process temperatures of 1800–2400 °C, volatile species of Si₂C, SiC₂ and Si evaporate from the source and deposit on the seed. The source temperature de-

termines the rate of evaporation, and the temperature difference between the source and the seed determines the diffusion transport rate; together these two temperatures govern the growth rate. The Si:C ratio is a key control parameter for proper growth of the desired material. Sublimation is normally performed in vacuum or inert gas, such as argon. The pressure of this gas also controls the growth rate to an extent. Growth rates can reach up to 0.5–1.0 mm/h, and crystal lengths of up to 20–40 mm are achieved. Doping can be accomplished via nitrogen and aluminium addition for n- and p-type material, respectively.

Polytype control is obtained by carefully choosing the seed orientation, with growth on the (0001)Si face giving 6H material, while 4H material grows on the (0001) face. Both growth temperature and pressure and intentional/unintentional impurities can also affect the polytype obtained. One remaining problem is that of micropipes within the material. These are hollow core defects of 1–10 μm diameter, although their density has been reduced from 1000 to 1 cm^{-2} in the best recent samples. Maintaining good control over the nucleation conditions can reduce micropipe density.

Diamond

High-pressure diamond is produced under conditions where it is thermodynamically more stable than graphite [12.64]. However, very high temperatures and pressures are required for growth, unless catalysts are also employed. Clearly a detailed knowledge of the diamond–graphite pressure–temperature phase diagram is a prerequisite for successful growth. For example, in static compression, 8–20 GPa of pressure and temperatures of 1000–3000 $^{\circ}\text{C}$ are needed. Carbon sources include graphite, amorphous carbon, glassy carbon and C_{60} . Dynamic compression techniques employ pressures of 7–150 GPa. A wide variety of catalysts have been used, ranging from conventional ones, such as transition metals, to carbide-forming elements (including Ti and Zr), to Mg, to oxygen-containing materials (carbonates and hydroxides for instance), to inert elements (such as P, Cu, Zn) to hydrides.

Crystals above 1 mm in size are grown by the temperature gradient method. Growth temperature, temperature gradient, growing time, type of catalyst and impurities all affect the growth. Crystals up to 10 mm in size are commercially available. Growth rates can reach 10 mg/h and are governed by the growth temperature and the temperature gradient. If a high-quality diamond seed is used, crystals free of major defects (such as inclusions, stacking faults and dislocations) can be obtained. Impurities,

mainly from the catalysts used, are still an issue, as is their nonuniform distribution.

12.3.2 Groups III–V

Gallium Arsenide (GaAs)

Mullin [12.44] provides a comprehensive discussion on all aspects of the growth of III–V and II–VI compounds. He starts his treatment with the early work done on purification in Ge and Si, which highlighted the need for both high purity and for single-crystal growth. He notes that *Pfann* [12.65] initiated zone melting, where, in the simplest case, a horizontal molten charge is progressively frozen from one end. Both zone leveling (in which a liquid zone within a solid bar is moved through the bar in one direction and then in the reverse direction, thus producing a uniform distribution of a dopant) and zone melting (in which liquid zones are repeatedly passed through a solid bar to force impurities to segregate to the ends of the bar, and then these are subsequently removed before the bar is used in growth) are used. Horizontal techniques were also used to produce single-crystal material by slight back-melting of a seed crystal, followed by progressive freezing. These various horizontal techniques were subsequently applied to the purification of both elements and compounds in materials from Groups III–V and II–VI.

Table 12.3, taken from [12.44], summarizes the growth techniques that have been used for the various compounds in Groups III–V and II–VI. There is much current debate about the relative advantages and disadvantages of these techniques as applied to compound crystal growth.

GaAs is the most important Group III–V compound. Horizontal growth can be used to produce low temperature gradients at the solid/liquid interface, which reduces stress-induced slip and hence dislocations. This is particularly beneficial for the use of GaAs in laser diodes, where dislocation densities are required to be very low. However, there is a potential problem in using these low temperature gradients – constitutional supercooling – which may result in unstable growth and second-phase inclusions. In addition to this, there is the issue of impurity uptake from the crucible and nucleation of twins, grain boundaries and even polycrystalline growth from the crucible walls. These problems can be minimized by taking care over experimental details, but they still make horizontal growth unattractive where large-area uniformity is required, particularly as scaling up to ever-larger diameters is difficult.

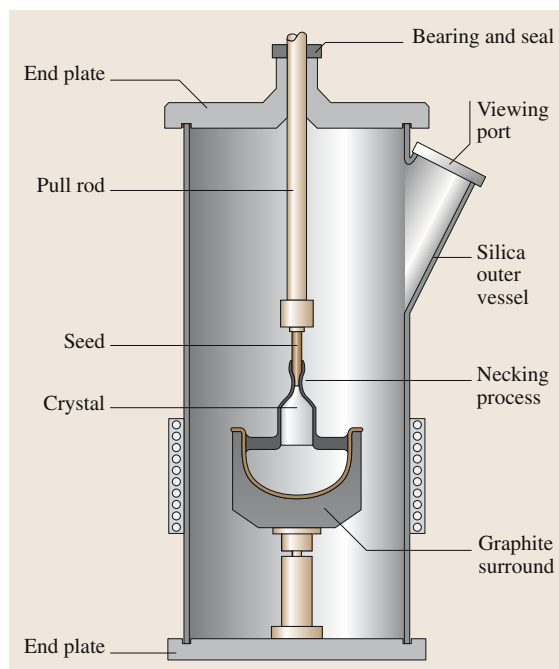
Table 12.3 General applicability of growth techniques to compound semiconductors. The more asterisks, the more appropriate the technique. *P*: potentially applicable; *C*: conventional VGF; *L*: LEC VGF. (From [12.44])

Technique/ Compound	Zone melting, horizontal Bridgman	VGF, vertical Bridgman	Conventional vertical pulling	Liquid encapsulation pulling	Vapor growth
InSb	***	P	***		P
GaSb	***	P	***		P
InAs	***	P		***	P
GaAs	***	C***; L***		***	P
InP	*	C*; L**		***	P
GaP	*	C*; L**		***	P
HgSe		**			P
HgTe		***			P
CdSe	*	**			***
CdTe	***	***			***
ZnSe		**			***
ZnTe		**			***
HgS					***
ZnS					***
CdS					***

Early efforts to produce this material by the vertical Czochralski method failed due to dissociation of the melt. This problem was solved by the LEC technique [12.44]. A layer of boric oxide floats on the melt and the seed crystal is dipped through this to contact the melt (Fig. 12.13). PBN crucibles are normally used and heating can be by resistance or RF means, but commercial systems use graphite heaters. Most semi-insulating GaAs, which is required for integrated circuits, is produced at pressures of 100–200 atm. The advantage of this modification is that elemental As and Ga can be used as the source. Diameter control is normally via crystal weighing. Currently, LEC produces material 150 mm in diameter [12.66] from 28–40 kg melts using 400 mm-diameter PBN crucibles. Also reported were modifications to the basic Czochralski technique to reduce stresses, and hence dislocations, in crystals up to 100–150 mm in diameter. These are

vapor pressure controlled Czochralski, fully encapsulated Czochralski and hot-wall Czochralski. The reader is referred to the reference for more details of these techniques.

Fig. 12.13 Schematic of liquid-encapsulated Czochralski growth equipment. A silica outer vessel with viewing port is held between end-plates. The induction heating coils couple into the graphite surround. The seed is fixed in the chuck on the pull rod that rotates and moves through the bearing and seal. The crystal grows from the seed through a necking process, and on withdrawal pulls out a layer of B_2O_3 over its surface. (After [12.44]; copyright Wiley-VCH (1998), used with permission)



An alternative to Czochralski is that of vertical gradient freeze. For reproducible growth of large-diameter crystals by this technique, it was found necessary to use B_2O_3 in a PBN crucible. Freezing is accomplished by moving the temperature gradient via furnace controller changes, rather than movement of the furnace itself. This naturally produces low temperature gradients, which give low dislocation densities, and it produces a crystal of the right size and shape for subsequent slicing and processing. *Rudolph* [12.66] reports that in 1999 LEC growth of GaAs accounted for $\approx 90\%$ of all SI GaAs, with the remaining 10% produced via vertical Bridgman or VGF growth. By 2000, he puts the figures at $\approx 50\%$ LEC and $\approx 50\%$ VGF. *Rudolph* also notes that VGF has been reported to grow material up to 150 mm in diameter.

Indium and Gallium Phosphides

As for GaAs, high-pressure LEC growth is normally used for both InP and GaP [12.44]. The problems involved are analogous to those for GaAs growth, with the major addition of an increased tendency for twinning. Loss of phosphorus (P_4) is still a problem and leads to a deterioration in crystal quality. Dislocation densities are still seen to be high. *Asahi* et al. [12.67] report that some of these problems have been alleviated by modifications to the basic LEC process, such as thermal-baffled LEC, phosphorus vapor pressure controlled LEC, and vapor pressure controlled Czochralski (VCZ). Dislocations tend to increase with crystal diameter, although crystals up to 100 mm in size with low dislocation densities have been grown recently by VCZ. Lower dislocation densities are produced by the VGF method at a given crystal diameter. High-pressure VGF growth of InP is difficult and temperature fluctuations cause twinning. By improving the temperature control (to within $\pm 0.03^\circ\text{C}$) *Asahi* et al. have shown that twin-free (100) InP single crystals 100 mm in diameter and 80 mm in length can be obtained. Temperature gradients of $< 10^\circ\text{C}/\text{cm}$ reduced the etch pit densities to below those of comparable LEC material.

Indium and Gallium Antimonides

Mullin [12.44] has detailed the early work on InSb; much less work has been done on GaSb. The low melting point combined with the negligible vapor pressure of Sb over the melt make zone-refining methods attractive. However, certain impurities render this route troublesome, although ways of overcoming this issue were found. Crystal-pulling methods were preferred for their versatility, although twinning was

seen as a major issue. Growth on the (111)Sb face produces this twinning and it is recommended that growth should occur on the [211]Sb or [311]Sb faces instead.

Micklethwaite and *Johnson* [12.68] have summarized the current status of the production of InSb for IR detector applications. InSb expands by $\approx 13\%$ on freezing, leading to a net flow away from the growth interface. This ‘density effusion’ must be accounted for when modeling this system. Hydrogen is used as the atmosphere, as it reduces floating InO_x . The basic Czochralski method is used for commercial production and diameters of 75 mm are routine. Defect densities can be as low as 10 cm^{-2} . Many of the details of the growth process are proprietary. Growth proceeds without a liquid encapsulant, and as such resembles Si growth more closely than its arsenide and phosphide cousins. However, there are significant differences from the silicon case in terms of reduced radiative heat transfer and increased convective flow in the hydrogen atmosphere. Fluid mechanics in the melt are complicated by both ‘density effusion’ and the facet effect [12.44].

Bridgman growth is not used for InSb growth any longer, but float-zone growth has been achieved in microgravity conditions using an oxide ‘skin’ for containment.

Group III Nitrides

There has been a rapid increase in the interest in these compounds following the successful development of blue laser diodes. Ideally, epitaxial thin films of these nitrides would be grown on bulk substrates of similar compounds, to minimize lattice mismatch. However, growth of these compounds by the more normal Czochralski and Bridgman techniques is not possible due to the high melting temperatures (2200–3500 K) and high decomposition pressures (0.2–60 kbar) [12.69, 70]. One growth method that can be used is high-pressure solution growth. In this technique, liquid Ga (plus 0.2–0.5 at.% Mg, Ca, Be or Zn) is held in a high-pressure chamber with a N_2 atmosphere. The maximum pressure is ≈ 20 kbar and the maximum temperature is 2000 K. Both pressure and temperature are controlled within tight limits and in situ annealing in vacuum is also allowed for. A temperature gradient of $2\text{--}20^\circ\text{C}/\text{cm}$ is maintained along the crucible axis and the N_2 dissolves in the hotter end and GaN crystallizes in the cooler end. No seeding was used and after 120–150 hours of growth hexagonal needles or platelets are produced of size 10×10 mm. Growth by MOVPE and MBE of GaN and InGaN MQWs on these bulk GaN samples was also re-

ported by Grzegory et al. [12.70]. LEDs and laser diodes based on these epitaxial layers were also described.

This technique only produces very small crystals of both AlN and InN.

Nishino and Sakai [12.71] describe the sublimation growth of both GaN and AlN. Initially, the source powder is synthesized by heating Ga metal in NH_3 . This powder is then annealed in NH_3 before loading into the growth chamber. Temperatures are 1000–1100 °C but atmospheric pressure can now be employed. Both NH_3 and N_2 are introduced into the growth chamber. Either small (3 mm \times several hundred μm) ‘bulk’ crystals or free-standing films (10–30 μm thick) on sapphire substrates are produced. Thicker samples (up to 500 μm thick) can be produced by the sublimation sandwich method in which the distance from the source to substrate is reduced to 2–5 mm. A hot-pressed polycrystalline source of AlN and a SiC-coated graphite crucible was used to grow 0.3–1.0 mm-thick films on the 6H-SiC seed (10 \times 10 mm in size) at temperatures of 1950–2250 °C.

12.3.3 Groups II–VI

Several recent reviews of the growth of a wide range of binary and ternary II–VI compounds have been published [12.44, 58, 72, 73]. One of the key concepts in the growth of these compounds is the ease with which phase transitions occur. This imposes limits on certain growth techniques for particular compounds.

There are two main types of bulk growth technique, namely growth from the liquid and growth from the vapor. Most narrow-gap II–VI compounds are characterized by high melting points and/or high component partial pressures, and early work concentrated on various forms of vapor growth, particularly for S- and Se-containing compounds. Problems with low growth rates and/or small crystal sizes and purities led to a switch to melt growth techniques, although there have been some recent significant developments in the growth of larger crystals from the vapor.

Many techniques have been used to grow narrow-gap II–VI compounds. Solid state recrystallization (SSR) is used for ternary systems where there is a wide separation between solidus and liquidus, such as MCT. Other names that have been used for this process are quench anneal (QA) and cast recrystallize anneal (CRA). The term anneal is used in the first case to define a high-temperature grain growth process, while in CRA it is a low-temperature process used to adjust stoichiometry. Strictly speaking, SSR is crystal growth from the

solid phase at temperatures close to the melting point, but it is included here for convenience. In the basic technique, pure elements are loaded into a cleaned silica ampoule and the charge is melted and rocked to ensure complete mixing. Charges are then normally quenched rapidly, into air or oil, to room temperature in the vertical orientation. This produces a dendritic structure that is reduced/removed by the recrystallization step, which proceeds at temperatures just below the melting point for many days. Grain growth occurs and micro-inhomogeneities in composition are removed. Care must be taken in the quenching stage to avoid pipes/voids that cannot be removed by the recrystallization step.

Bridgman growth is the most widely used of the bulk growth techniques applied to narrow-gap II–VI compounds. It is the only technique that has produced crystals of all the binary and ternary compounds studied. Numerous modifications have been applied to the basic process, but the three principal means of achieving growth are to move an ampoule through a temperature gradient, to move the furnace past a stationary ampoule and to move a temperature profile from high to low temperatures with both the furnace and ampoule stationary, the so-called vertical gradient freeze (VGF) method.

Silica ampoules are usually used but various coatings have been applied, and the use of other materials such as PBN has been reported [12.74]. Charges are prepared either using the appropriate elements or, for ternaries and quaternaries, preformed compounds of the binaries. Both vertical and horizontal configurations have been used and large crystal sizes coupled with a relatively high growth rate (1–2 mm/h) make the basic technique relatively cheap and versatile. Marked segregation of both matrix elements in ternaries and quaternaries and impurities can occur. Difficulties with controlling component vapor pressures can be overcome by employing a reservoir of one of the components at a temperature lower than that of the growing crystal. High partial pressures in Hg-containing compounds can cause ampoule failure and ways have been developed to cope with this situation [12.75].

The traveling heater method (THM) has mainly been used for Te-based binary and ternary compounds. Growth by THM combines the advantages of low-temperature solution growth with steady-state conditions, as in zone melting. A homogeneous alloy can be used as the starting ingot, or segments of the binaries can be employed. If the Te zone height is made equal to the ring-heater size, then a flat interface is obtained for particular geometries of crystal and furnace and growth parameters [12.76]. Very low growth rates (0.1 mm/h)

are typical, but diameters of up to 40 mm have been accomplished. As in Bridgman, many variations on the basic THM method have been attempted. These include multipass THM [12.77] and ‘cold’ THM [12.78].

Hydrothermal growth has mainly been applied to S- and Se-based binary compounds and has been reviewed by Kuznetsov [12.79]. The main advantages are that it reduces growth temperature (to 200–250 °C) and that it deals with the high partial pressures. Growth rates are very low (0.05 mm/d), and crystals are small (few mm) and contain subgrains and inclusions, but can still provide useful fundamental data.

Growth from the vapor is usually carried out at much lower temperatures than growth from the liquid, below unwanted phase transition temperatures, and can lead to less defective crystals. However, growth rates are generally much lower and uncontrolled nucleation and twinning are more prevalent. Techniques can be split essentially into unseeded and seeded methods. The simplest technique is that of chemical vapor transport (CVT), usually using iodine as the transport agent. The compound to be grown is reacted at a high temperature with the iodine, transported to a cooler region and deposited. The reverse reaction produces iodine, which diffuses back to the high-temperature region and the process repeats itself. The method uses a two-zone furnace but has the disadvantages of iodine incorporation and small crystal sizes. Mimila and Triboulet [12.80] have experimented with sublimation and CVT using water as a transporting agent for ZnSe, and Triboulet (private communication, 1996) has succeeded in growing CdTe using this method.

Most other vapor techniques are in some sense a derivative of the Piper–Polich method (Fig. 12.12b). Although this was developed originally for the growth of CdS [12.81], the technique may be applied to the vapor growth of any compound that sublimates readily below the melting point temperature. The crystal is grown in a closed crucible that is moved through a steep temperature gradient, such that the source material is always hotter than the growing surface and mass transport occurs from source to crystal.

Vapor growth under the controlled partial pressure of one of the constituents has been obtained by including small quantities of one of the elements in an extended part of the capsule located in a cooler part of the furnace [12.82]. The technique, termed ‘Durham’, is a vertical unseeded growth procedure, where the source is sublimed from bottom to top in an evacuated silica capsule over several days. Constituent partial pressures are independently controlled by placing one of the ele-

ments in a separately heated reservoir connected to the main capsule via a small orifice.

The vertical unseeded vapor growth (VUVG) method differs from the Durham technique in that the partial pressures are controlled by the initial deviation from stoichiometry of the charge material. It was initially developed to grow CdTe. The crystal is grown in an evacuated cylindrical silica capsule in a vertical furnace. A long quartz tube, which extends outside the furnace, is attached to the top of the capsule, providing a heat pipe to create a cooler area for nucleation, which is closely monitored to ensure that the nucleus is single-grained. If it is not, then the capsule is returned to the starting position and the process repeated.

The sublimation traveling heater method (STHM) was developed from THM by Triboulet and Marfaing [12.77]. The Te-rich molten zone in conventional THM is replaced by an empty space or vapor zone. A problem arises in that the constituent partial pressures are not only functions of the temperature. However, the vapor can be maintained near the stoichiometric P_{\min} condition if a small capillary, one end of which is at room temperature, is connected to the vapor chamber [12.83]. Excess species are sublimed down the capillary, preserving the relative constituent partial pressures and hence the growth rate.

In one version of the seeded technique, an oriented seed is placed on top of a long sapphire rod inside a sealed quartz capsule. The source material is placed in a basket above the seed and sublimation is top-down. The sapphire rod is carefully centered within the growth capsule to ensure a narrow gap between it and the walls of the capsule. Crystals are grown in vacuum or in about 200 Torr of Ar, H₂ or NH₄I. The crystals grow clear of the container walls and generally consist of two or three large grains, some of which maintain the seed orientation. This ‘Markov’ technique has been adapted for commercial production by the ELMA Research and Development Association of Moscow (cited in Durose et al. [12.84]). The seed is mounted on an optically heated sapphire rod heat pipe, in a continuously pumped chamber, which can be back-filled with He to control growth rate. Large crystals of CdTe, 50 mm in diameter and up to 10 mm in length have been grown.

It is clear that HgS is grown by vapor means or by hydrothermal techniques, both of which are at low temperatures and hence low S partial pressures. HgSe has been produced by vapor growth and also by the Bridgman technique. Thick-walled silica ampoules are required in the growth of the Hg-containing compounds to contain the high Hg partial pressures. More details on

the growth of Hg-based binaries can be found in [12.78] for vapor and Bridgman growth, while *Kuznetsov* has reviewed the hydrothermal growth of HgS [12.79].

Early work on the growth of the Cd-based binaries was summarized by *Lorenz* [12.85]. Vapor growth was favored for CdS and CdSe, although hydrothermal growth was also discussed. The latter technique, as applied to CdS, was discussed in [12.79]. CdTe was grown by various melt growth techniques, including seeded zone melting and Bridgman, and by vapor growth methods. *Zanio* [12.86] has given a review of the growth, properties and applications of CdTe, including radiation detectors. By that time, vapor growth had been essentially supplanted by both Bridgman and THM growth. A more recent review of CdTe can be found in *Capper* and *Brinkman* [12.87]. The main uses of CdTe (and CdZnTe/CdTeSe) are as a substrate material for the epitaxial growth of MCT and, more recently, as radiation detectors [12.88]. *Triboulet* et al. [12.89,90] applied SSR to the growth of both CdSe and CdTe. The prime reasons for the use of this technique were the high melting point of CdSe and the solid-state phase transitions in the case of CdTe.

A wide variety of growth techniques from the liquid have been used to produce crystals of CdTe. These include solvent evaporation [12.91], liquid encapsulated Czochralski [12.92], zone refining [12.93], VGF [12.94,95], solution growth [12.96], heat exchanger method [12.97], float-zoning [12.98], Bridgman [12.87] and THM [12.47], and *Triboulet* et al. [12.89] have reviewed these materials as substrates for MCT epitaxial growth and compared them to the alternatives based on GaAs, sapphire and Si. Problems with the growth of CdTe include low thermal conductivity, difficulty in seeding due to the need for superheated melts, ease of twin formation, tilts and rotations in the lattice, stoichiometry control and impurities. However, *Triboulet* et al. [12.90] concluded that despite the progress made in the alternative substrate materials, lattice-matched substrates based on CdTe produced MCT epitaxial layers that gave the best device performance.

Within the Bridgman growth technique, many alternatives have been attempted. These include holding excess Cd at a lower temperature in a separate reservoir [12.99], vibration-free growth [12.100], vibrational stirring [12.101], horizontal growth [12.102], low temperature gradients [12.103], high-pressure (100 atm) growth [12.104] and the addition of the accelerated crucible rotation technique (ACRT) [12.105]. Work in the US [12.102] aimed at improving both the horizontal and

vertical techniques and scaling-up to large melts, up to 8 kg (horizontal). Recent work by *Szeles* et al. [12.106] has shown that crystals up to 140 mm in diameter and 10 kg in weight can be grown by careful control of the temperature fields.

In the Bridgman process, elemental Cd and Te are loaded into a clean carbon-coated silica ampoule, homogenized by melting/rocking, heated to temperatures in excess of 1100 °C, and then frozen at rates of a few mm/h in a vertical or horizontal system. Single crystals or, more normally, large-grained ingots of size ≈ 75 mm in diameter and 10–15 cm in length are produced. Seeding on the (111) is often utilized but it is not clear that it results in better growth in every case.

Growth by variations on the basic THM process have included the addition of ACRT [12.107,108], focused radiant heating [12.109], sublimation THM and multipass THM [12.77] and ‘cold’ THM at 780 °C [12.78]. Crystals are grown either from the pre-compounded binaries or from the elements (as in cold THM). Growth rates tend to be low, a few mm/d, but the crystals are very pure due to segregation effects and high-resistivity material results ($10^7 \Omega \text{ cm}$ in [12.77]). This makes THM material ideally suited for gamma and X-ray detectors.

The vast majority of work on bulk growth of MCT has been from the melt. Although rapid progress has taken place in epitaxial growth techniques for MCT, material grown by several bulk methods is still in use for infra-red detection, particularly for photoconductive detectors (this is the case in this author’s lab and various others worldwide, such as AIM in Germany). Several historical reviews of the development of bulk MCT have been published [12.110–112]. Many techniques were tried in the early years but three prime techniques survived: SSR, Bridgman and THM.

Two fundamentally different approaches have been followed to improve the basic Bridgman process. These are based on controlling melt mixing and heat flows, respectively. The former has been studied by this author and coworkers [12.113] while the latter includes the work of *Szofran* and *Lehoczky* [12.114], among others. In the Bridgman process, elemental Cd, Hg and Te are loaded into a clean silica ampoule, homogenized by melting/rocking and then frozen slowly from one end in a vertical system to produce a single crystal or, more normally, a largegrained ingot. Marked segregation of CdTe with respect to HgTe occurs in the axial direction, but this leads to an advantage of the Bridgman process over other techniques: material in both ranges of interest (3–5 and 8–12 μm , for $x = 0.3$ and 0.2, respectively) is produced in a single run.

A means of stirring melts contained in sealed, pressurized ampoules was needed and the **ACRT** of *Scheel* and *Schulz-Dubois* [12.41], in which the melt is subjected to periodic acceleration/deceleration at rotation rates of up to 60 rpm, was chosen. The first report of **ACRT** in **MCT** Bridgman growth was given in a patent [12.115]. These effects were developed and discussed in more detail in later papers, which are reviewed in [12.113]. Crystals were produced up to 20 mm in diameter and with x values of up to 0.6 in the tip regions of some crystals (this has recently been increased to ≈ 0.7 at 20 mm diameter in the author's laboratory [12.42, 116]).

Triboulet [12.47] developed this technique for **MCT**, where diameters of up to 40 mm have been accomplished as well as x values of up to 0.7 [12.117] for optical communication devices. *Durand* et al. [12.118] employed seeds to produce large, oriented crystals, an advantage of this technique over other bulk methods. *Gille* et al. [12.119] adopted a slightly different approach. A pre-**THM** step is carried out to quench a Te-rich (53–60%) **MCT** melt. A first **THM** run at 2 mm/d is then used to provide the source ingot, after removal of the tip and tail sections, for the final **THM** run. The entire growth procedure takes several months but gives uniform material. This group has also used rotation in the horizontal growth by **THM** with some success [12.120]. Attempts to use 'cold' **THM** (starting from the elements) led to little success for **MCT** nor did use of a Hg reservoir to control the process. Two groups have applied **ACRT** to **THM** and obtained enhanced material properties. *Royer* et al. [12.121] used a saw-tooth **ACRT** sequence and obtained improved radial and axial compositional uniformity. *Bloedner* and *Gille* [12.122] used **ACRT** although no significant dependence of crystallinity on rotation sequence was seen. They achieved an increase in growth rate from 1.5 to 8.5 mm/d.

Bridgman is the main melt growth technique for the Hg- and Cd-based ternary and quaternary compounds, while work on the vapor growth of these compounds has been limited. The various Hg-based ternaries (such as HgZnTe and HgMnTe) have been studied as potential alternatives to **MCT** for infrared detectors (*Rogalski* [12.123] has more details including phase equilibria plots). This work was initiated by a theoretical prediction that the Hg–Te bond is stabilized by the addition of Zn, in particular. *Rogalski* concluded that **THM** produced the best quality HgZnTe while Bridgman growth of HgMnTe was described as being similar to **MCT** growth but with reduced segregation. Crystals of HgMnTe up to 40 mm in diameter were produced.

Triboulet [12.124] has discussed alternatives to **MCT**. He concluded that HgMgTe, HgCdSe and HgZnSe are not suitable, for a variety of reasons, but that HgMnTe and HgZnTe are potentially suitable.

The main Cd-based ternaries are CdZnTe and CdTeSe, which are used as substrates for epitaxial **MCT** growth. A vast literature exists on these compounds and it is still a very active area of research. Vertical and horizontal Bridgman are the main techniques used [12.87], although **VGF** has been used [12.87], as has **THM** [12.47]. High-quality crystals up to 100 mm (and even 125 mm) in diameter can be produced by **VGF** [12.67, 95]. Even larger crystals are being grown by horizontal Bridgman [12.106, 125]. Two main problems remain in these compounds. These are the uniformity of Zn or Se, and impurities.

Recent reports by *Pellicciari* et al. [12.126] have shown how solvent evaporation from a Te-rich solution of CdTe in an open tube system can produce large-grained and even single crystals of CdTe up to 300 mm in diameter. This large-area material is aimed at the fabrication of X-ray and γ -ray detectors.

Dilute magnetic semiconductors (**DMSs**) are a class of semiconductor where the semiconducting properties are changed through the addition of a magnetic ion (such as Mn^{2+} or Fe^{2+}). These materials therefore display normal electrical and magnetic properties, and novel ones such as large Faraday rotations and giant magnetoresistance. *Pajaczowska* [12.127] gave an early review of the phase diagrams, lattice parameters and growth of many of the Mn- and Fe-based **DMS** compounds. Vapor growth and hydrothermal growth were mainly used, although the Te-based compounds were being grown by the Bridgman technique. Only small additions of the magnetic ions are required, so essentially the same techniques can be used as those applied to the parent binaries. The most comprehensive review of **DMSs** is given in [12.128]. In that book, *Giriat* and *Furdyna* review the crystal structures of, and growth methods used for, the Mn-based compounds. The majority of the compounds, particularly the Te-based ones, were grown using Bridgman or unspecified modifications of it. CdMnS was grown in graphite crucibles under high inert gas pressure, or by chemical vapor transport (using iodine as the transport agent). HgMnS was grown hydrothermally, yielding 2 mm-diameter crystals. Similar preparation methods to those used for the Mn-based compounds were applied for Fe-based compounds, but less Fe can be incorporated in the lattice.

Bismuth telluride (strictly speaking a V–VI compound) is used in thermoelectric applications close to

room temperature [12.129]. This material is also grown by the Bridgman technique. Both n-type and p-type materials are required to make the thermoelectric devices, and these can be achieved either by stoichiometric deviations or deliberate doping. Typical growth rates are 0.5–2.9 mm/h for 8 mm-diameter ingots of up to 85 mm length. In this author's lab, bismuth telluride is grown at similar rates for 12 mm-diameter ingots up to 100 mm length using antimony and iodine as acceptor and donor dopants, respectively.

12.3.4 Oxides/Halides/Phosphates/Borates

Oxides

Hurle [12.43] summarized issues concerning the growth of oxides by the Czochralski technique. The high melting points of most oxides of interest necessitate the use of precious metal crucibles, such as iridium or platinum, although molybdenum and tungsten ones can be used for lower melting point materials. Extensive and efficient thermal insulation is required, using alumina, magnesia, zirconia and thoria. Some materials require a partial pressure of oxygen in the growth chamber to prevent reduction of the melt. However, this can lead to oxidation of the crucible, which in turn enters the melt. Careful management of the oxygen pressure and gas flow is needed, together with physical baffles. Heating is normally by means of RF. Mounting of the seeds is problematic and long seeds are necessary to avoid thermal degradation of the seed holder and pull rod. Rotation or translation of the crucible is not normally used so the rate of growth has to be above the pulling rate to account for the fall in melt height. Crystals are normally limited to less than half the crucible diameter for stable growth. Growth rates are low, a few mm/h or less, to avoid unstable growth. Afterheaters are also employed to reduce the built-in stress in the crystal during the later stages of growth and subsequent cool-down. Growth runs extend over several days, so automatic diameter control, in this instance using weighing techniques, is mandatory. The high temperatures lead to high temperature gradients and thus strong buoyancy-driven convection that is turbulent by nature. This turbulence gives rise to nonuniformities in the crystals. In addition, catastrophic failure can result due to the coupling between this turbulent flow and the flow due to the crystal rotation. Thus, an understanding of the convective flows present during oxide growth is critical if large-diameter highly perfect crystals are to be produced.

More recently, *Fukuda* et al. [12.130, 131] have presented detailed reviews of the growth of a wide

range of oxide materials of current interest for optical and optoelectronic applications. They firstly note that only LiNbO_3 (LN), LiTaO_3 (LT) and $\text{Y}_3\text{Al}_5\text{O}_{12}$ (YAG) have as yet reached production status. Many of the other materials exhibit superior properties but are not available in large quantities at sufficient perfection to warrant them being used in devices. Recently, crystals of new nonlinear optical materials, piezoelectric materials and new laser materials have attracted growing interest. In general, the growth methods are based on liquid–solid transitions, either from the melt or from high-temperature solutions (flux), for congruently and incongruently melting materials, respectively. Czochralski growth is still the mainstay of oxide growth, although problems with shape (diameter) control of some oxides (including rare earth vanadates and rutile) necessitate the use of edge-defined film-fed growth (EFG). Growth from fluxes is used for a range of oxide materials, although some incongruently melting materials can be produced from melts of the same composition as the crystal by a micro-pulling down technique.

In the micro-pulling down technique, the melt is held in a Pt crucible and a micronozzle (die) is arranged at the bottom of the melt. An after heater is placed below this arrangement. This combination produces a steep temperature gradient (300 °C/mm), which ensures stable growth at high rates. Single crystals of 0.05–1.0 mm diameter are produced onto seeds. LN crystals, for example, are grown at 12–90 mm/h from stoichiometric melts up to 100 mm in length, assumed to be limited only by the melt volume used. Dislocations are introduced if the diameter of the crystal exceeds 0.8 mm. Growth of $\text{K}_3\text{Li}_2\text{Nb}_5\text{O}_{12}$ (KLN), which has outstanding electro-optic and nonlinear optical properties, is problematic by either Czochralski or Kyropoulos methods due to segregation effects at the solid/liquid interface and cracking. This material was grown by the micro-pulling down technique at rates of 20–80 mm/h with diameters of 0.15–0.5 mm. The crystal composition was found to be near to the melt composition and crystals were single-domain.

Crystals in the KTiOPO_4 (KTPO) family were grown by the flux method using slow-cooling from 1100 to 800 °C. Crystals a few millimeters in size were obtained for second harmonic-generation measurements.

Crystals from the langasite family ($\text{La}_3\text{Ga}_3\text{SiO}_{12}$) are used for lasers. These were grown by the Czochralski method to 50 mm in diameter and 130 mm in length. High-purity oxide powders were used as starting materials and an Ar/O_2 mixture was used to suppress evaporation of the gallium suboxide. A pulling rate of

1.0 mm/h was used together with a crystal rotation rate of 10 rpm. The resulting crystals were single-phase and crack-free. These crystals show superior properties, as filters, to quartz. Crystals of up to 100 mm in diameter have been grown recently [12.132].

Both rare earth vanadates and some garnet crystals show a tendency toward spiral growth when the Czochralski method is used. This tendency can be circumvented by adjusting the temperature gradient in the melt.

Another way of overcoming problems with the Czochralski growth of these materials is to use the EFG method. Rutile (TiO_2) is used as a polarizer in optical isolators and prisms for optical communication systems. Crystals of this material can be grown by Verneuil and float-zone methods, but growth by Czochralski is difficult. Iridium is used for the die material and crystals up to 8×1 cm and 0.5 mm thick were grown. A modified die was used to grow a rod of 15 mm diameter and ≈ 40 mm in length. A double-die system was shown that was used to produce a core-doped crystal of LN, using Nd^{3+} and Cr^{3+} as the dopants. A 5 mm-diameter crystal some 60 mm in length was produced. Such crystals have potential in optical applications where pumping energy is only absorbed in the core region, reducing heat-induced energy losses.

Quartz is used in both piezoelectric and in various optical applications [12.133]. *Balitsky* [12.134] noted that some 2500 t are produced per year. Hydrothermal growth is the main growth technique used in order to keep the growth temperature below the β - (non-piezoelectric) to α -quartz (piezoelectric) phase transition. Growth at temperatures of 250–500 °C and pressures of 50–200 MPa are used in both alkaline and fluoride solutions. Growth rates of 0.5–1.0 mm/d are achieved on (0001) or (1120) seeds and crystals up to $200 \times 70 \times 60$ mm in size weighing from 1–18 kg.

Various oxide materials are used widely in scintillator applications. The most common are $\text{Bi}_4\text{Ge}_2\text{O}_{12}$ (BGO) and PbWO_4 (PWO) [12.135, 136]. These materials are grown by normal Czochralski and Bridgman techniques and the requirement is for thousands of each for large high-energy/astrophysics projects in several centers throughout the world. For example, some 85 000 PWO crystals are required for one application at the Large Hadron Collider at CERN. High-purity starting materials are important, as are certain growth parameters and post-growth annealing, for example in oxygen, to reduce stress in the material.

Sapphire is grown by several techniques for a wide variety of optical, microwave and microelectronics ap-

plications [12.137]. Verneuil [12.138], Czochralski and EFG [12.43] have been used extensively to produce large single crystals, particularly for laser rods, and shaped pieces using various forms of dies. In addition, the horizontal directional solidification crystallization (HDC) technique [12.137] has been used as it is a simple and inexpensive technique to set up and use. A molybdenum boat filled with the charge is simply moved through the temperature gradient. Low temperature gradients lead to crystals with low stress and low dislocation densities (compared with the Verneuil and pulling techniques. The growth rate is high at 8 mm/h, and crystals of up to 10 kg and up to $30 \times 30 \times 2.5$ cm are produced. Even larger crystals, up to 20 kg with a diameter of 200 mm, can be produced by a modified Kyropolous method [12.139].

Another technique used is the heat exchanger method, which is a solidification technique from the melt [12.140]. Large crystals can be grown onto seeds by independently controlling the heat input and the heat output. The seed is prevented from melting by a helium gas flow at the bottom of the melt. By progressively increasing the helium flow, following partial melting of the seed, and/or decreasing the furnace temperature, growth is progressed. After complete solidification, an annealing step is performed prior to cooling to room temperature. The submerged solid/liquid interface damps out any thermal and mechanical variations and also produces a low temperature gradient at the interface. There is impurity segregation into the last-to-freeze parts of the crystal that form near the crucible walls. These areas are removed before machining takes place. Crystals up to 340 mm in diameter (65 kg) have been produced in this way. Slabs cut from these crystals are used as optical windows, between 5–25 mm thick.

Halides

Several halides form another class of scintillator materials. These are used in a range of applications, from high-energy physics and nuclear medicine to environmental monitoring and security systems [12.141]. Both large diameters (comparable with the human body) and large numbers are required for high-energy physics applications. The simple Bridgman growth process is still used for many of these materials, although problems do arise due to the contact between crystal and crucible. Insufficient melt mixing also leads to inclusions and striations in crystals. The alternative technique for growth is that of continuously fed Czochralski/Kyropolous growth. A melt-level sensor is normally used to control the diameter over the tens/hundreds of hours of growth needed for large halide crystals. Replenishment

of the melt by molten material from a closed feeder enables the starting materials to be purified, and a conical crucible that is not too deep is used. Crystals of CsI up to 500 mm in diameter and 750 mm in height weighing ≈ 550 kg have been produced [12.142]. The ultimate aim is to produce crystals > 1 m in diameter so that entire crystals can be used to monitor the human body.

Phosphates and Borates

Nonlinear optical materials are very important for laser frequency conversion applications. One of the most important of the phosphates is potassium dihydrogen phosphate (KDP), which is used for higher harmonic generation in large laser systems for fusion experiments [12.94, 143]. Growth takes place at room

temperature to 60°C , and growth rates can be as high as 10–20 mm/day, with sizes of $\approx 40\text{ cm} \times 40\text{ cm} \times 85\text{ cm}$ [12.143] or $45\text{ cm} \times 45\text{ cm} \times 70\text{ cm}$ [12.94], the latter quoted as taking over a year to grow!

Another important phosphate is potassium titanyl phosphate (KTP), used to obtain green light by frequency doubling a Nd:YAG laser. Growth in this case is from high-temperature solution at about 950°C [12.94]. Sizes of up to $32\text{ mm} \times 42\text{ mm} \times 87\text{ mm}$ (weight 173 g) can be grown in 40 days.

Borates, including barium borate, lithium borate, cesium borate and coborates such as cesium lithium borate are used in UV-generation applications. Crystals are again grown by the high-temperature solution method up to $14\text{ cm} \times 11\text{ cm} \times 11\text{ cm}$ in size, weighing 1.8 kg, in 3 weeks [12.94].

12.4 Conclusions

This chapter has summarized the current status of the bulk growth of crystals for optoelectronic and electronic applications. It is not intended to be a completely comprehensive view of the field, merely serving to introduce the reader to the wide range of materials produced and the numerous crystal growth techniques that have been developed to grow single crystals. An historical perspective has been attempted to give the

reader a feel for the scale of some of the activities. The sections on specific materials try to summarize the particular growth techniques employed, and those that cannot in some cases, and outline the typical sizes currently produced in the commercial and R&D sectors. For more details on current developments, the reader should refer to the books given in references [12.42, 60].

References

- 12.1 J. C. Brice: *Crystal Growth Processes* (Blackie, London 1986)
- 12.2 H. J. Scheel: *J. Cryst. Growth* **211**, 1 (2000)
- 12.3 H. E. Buckley: *Crystal Growth* (Wiley, New York 1951)
- 12.4 J. G. Burke: *Origins of the Science of Crystals* (Univ. California Press, Berkeley 1966)
- 12.5 D. Elwell, H. J. Scheel: *Crystal Growth from High-Temperature Solutions* (Academic, New York 1975)
- 12.6 H. J. Scheel: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 12.7 A. A. Chernov: *J. Mater. Sci. Mater. El.* **12**, 437 (2001)
- 12.8 A. V. L. Verneuil: *Compt. Rend. (Paris)* **135**, 791 (1902)
- 12.9 W. Nernst: *Z. Phys. Chem.* **47**, 52 (1904)
- 12.10 M. Volmer: *Z. Phys. Chem.* **102**, 267 (1927)
- 12.11 W. Kossel: *Nachr. Gesellsch. Wiss. Göttingen Math.-Phys. Kl.* 135 (1927)
- 12.12 I. N. Stranski: *Z. Phys. Chem.* **136**, 259 (1928)
- 12.13 G. Spezia: *Acad. Sci. Torino Atti* **30**, 254 (1905)
- 12.14 G. Spezia: *Acad. Sci. Torino Atti* **44**, 95 (1908)
- 12.15 J. Czocharlski: *Z. Phys. Chem.* **92**, 219 (1918)
- 12.16 S. Kyropoulos: *Z. Anorg. Chem.* **154**, 308 (1926)
- 12.17 P. W. Bridgman: *Proc. Am. Acad. Arts Sci.* **58**, 165 (1923)
- 12.18 P. W. Bridgman: *Proc. Am. Acad. Arts Sci.* **60**, 303 (1925)
- 12.19 F. Stöber: *Z. Kristallogr.* **61**, 299 (1925)
- 12.20 D. C. Stockbarger: *Rev. Sci. Instrum.* **7**, 133 (1936)
- 12.21 H. C. Ramsberger, E. H. Malvin: *J. Opt. Soc. Am.* **15**, 359 (1927)
- 12.22 G. K. Teal, J. B. Little: *Phys. Rev.* **78**, 647 (1950)
- 12.23 W. C. Dash: *J. Appl. Phys.* **30**, 459 (1959)
- 12.24 W. G. Pfann: *Trans. AIME* **194**, 747 (1952)
- 12.25 H. C. Theurer: US Patent, 3 060 123 (1952)
- 12.26 P. H. Keck, M. J. E. Golay: *Phys. Rev.* **89**, 1297 (1953)
- 12.27 F. C. Frank: *Discuss. Farad. Soc.* **5**, 48 (1949)

- 12.28 W. K. Burton, N. Cabrera, F. C. Frank: *Philos. Trans. A* **243**, 299 (1951)
- 12.29 G. P. Ivantsov: *Dokl. Akad. Nauk SSSR* **81**, 179 (1952)
- 12.30 G. P. Ivantsov: *Dokl. Akad. Nauk SSSR* **83**, 573 (1953)
- 12.31 W. A. Tiller, K. A. Jackson, J. W. Rutter, B. Chalmers: *Acta Metall. Mater.* **1**, 428 (1953)
- 12.32 A. E. Carlson: PhD Thesis, Univ. of Utah (1958)
- 12.33 H. J. Scheel, D. Elwell: *J. Cryst. Growth* **12**, 153 (1972)
- 12.34 J. A. Burton, R. C. Prim, W. P. Slichter: *J. Chem. Phys.* **21**, 1987 (1953)
- 12.35 W. van Erk: *J. Cryst. Growth* **57**, 71 (1982)
- 12.36 D. Rytz, H. J. Scheel: *J. Cryst. Growth* **59**, 468 (1982)
- 12.37 L. Wulff: *Z. Kristallogr. (Leipzig)* **11**, 120 (1886)
- 12.38 L. Wulff: *Z. Kristallogr. (Leipzig)* **100**, 51 (1886)
- 12.39 F. Krüger, W. Finke: *Kristallwachstumsvorrichtung*, Deutsches Reichspatent DRP 228 246 (5.11.1910) (1910)
- 12.40 A. Johnsen: *Wachstum und Auflösung der Kristalle* (Wilhelm Engelmann, Leipzig 1910)
- 12.41 H. J. Scheel, E. O. Schulz-Dubois: *J. Cryst. Growth* **8**, 304 (1971)
- 12.42 P. Capper (Ed.): *Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester 2005)
- 12.43 D. T. J. Hurle: *Crystal Pulling from the Melt* (Springer, Berlin, Heidelberg 1993)
- 12.44 J. B. Mullin: *Compound Semiconductor Devices: Structures and Processing*, ed. by K. A. Jackson (Wiley, Weinheim 1998)
- 12.45 C. J. Jones, P. Capper, J. J. Gosney, I. Kenworthy: *J. Cryst. Growth* **69**, 281 (1984)
- 12.46 P. Capper: *Prog. Cryst. Growth Ch.* **28**, 1 (1994)
- 12.47 R. Triboulet: *Prog. Cryst. Growth Ch.* **28**, 85 (1994)
- 12.48 M. Shiraishi, K. Takano, J. Matsubara, N. Iida, N. Machida, M. Kuramoto, H. Yamagishi: *J. Cryst. Growth* **229**, 17 (2001)
- 12.49 K. Hoshikawa, Huang Xinming, T. Taishi: *J. Cryst. Growth* **275**, 276 (2004)
- 12.50 L. Jensen: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.51 T. Ciszek: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 12.52 J. B. Mullin, D. Gazit, Y. Nemirovsky (Eds.): *J. Cryst. Growth* **189** (1999)
- 12.53 J. B. Mullin, D. Gazit, Y. Nemirovsky (Eds.): *J. Cryst. Growth* **199** (1999)
- 12.54 T. Hibiya, J. B. Mullin, W. Uwaha (Eds.): *J. Cryst. Growth* **237** (2002)
- 12.55 T. Hibiya, J. B. Mullin, W. Uwaha (Eds.): *J. Cryst. Growth* **239** (2002)
- 12.56 K. Nakajima, P. Capper, S. D. Durbin, S. Hiyamizu (Eds.): *J. Cryst. Growth* **229** (2001)
- 12.57 T. Duffar, M. Heuken, J. Villain (Eds.): *J. Cryst. Growth* **275** (2005)
- 12.58 P. Capper (Ed.): *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications* (Chapman Hall, London 1997)
- 12.59 P. Capper, C. T. Elliott (Eds.): *Infrared Detectors and Emitters: Materials and Devices* (Kluwer, Boston 2001)
- 12.60 H. J. Scheel, T. Fukuda (Eds.): *The Technology of Crystal Growth and Epitaxy* (Wiley, Chichester 2003)
- 12.61 S. Nishino: *Properties of Silicon Carbide*, EMIS Datarev. Ser. **13**, ed. by G. L. Harris (IEE, London 1995)
- 12.62 A. O. Konstantinov: *Properties of Silicon Carbide*, EMIS Datarev. Ser. **13**, ed. by G. L. Harris (IEE, London 1995)
- 12.63 N. Nordell: *Process Technology for Silicon Carbide Devices*, ed. by C. M. Zetterling (IEE, London 2002)
- 12.64 H. Kanda, T. Sekine: *Properties, Growth and Applications of Diamond*, EMIS Datarev. Ser. **26**, ed. by M. H. Nazare, A. J. Neves (IEE, London 2001)
- 12.65 W. G. Pfann: *Zone Melting*, 2nd edn. (Wiley, New York 1966)
- 12.66 P. Rudolph: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 12.67 T. Asahi, K. Kainosho, K. Kohiro, A. Noda, K. Sato, O. Oda: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 12.68 W. F. J. Micklethwaite, A. J. Johnson: *Infrared Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C. T. Elliott (Kluwer, Boston 2001)
- 12.69 I. Grzegory, S. Porowski: *Properties, Processing and Applications of Gallium Nitride and Related Semiconductors*, EMIS Datarev. Ser. **23**, ed. by J. H. Edgar, S. Strite, I. Akasaki, H. Amano, C. Wetzel (IEE, London 1999)
- 12.70 I. Grzegory, S. Krukowski, M. Leszczynski, P. Perlin, T. Suski, S. Porowski: *Nitride Semiconductors: Handbook on Materials and Devices*, ed. by P. Ruterana, M. Albrecht, J. Neugebauer (Wiley, Weinheim 2003)
- 12.71 K. Nishino, S. Sakai: *Properties, Processing and Applications of Gallium Nitride and Related Semiconductors*, EMIS Datarev. Ser. **23**, ed. by J. H. Edgar, S. Strite, I. Akasaki, H. Amano, C. Wetzel (IEE, London 1999)
- 12.72 P. Rudolph: *Recent Developments of Bulk Crystal Growth 1998*, ed. by M. Isshiki (Research Signpost, Trivandrum, India 1998) p. 127
- 12.73 H. Hartmann, K. Bottcher, D. Siche: *Recent Developments of Bulk Crystal Growth 1998*, ed. by M. Isshiki (Research Signpost, Trivandrum, India 1998) p. 165
- 12.74 B. J. Fitzpatrick, P. M. Harnack, S. Cherin: *Philips J. Res.* **41**, 452 (1986)
- 12.75 P. Capper, J. E. Harris, D. Nicholson, D. Cole: *J. Cryst. Growth* **46**, 575 (1979)

- 12.76 R. Triboulet, T. Nguyen Duy, A. Durand: J. Vac. Sci. Technol. A **3**, 95 (1985)
- 12.77 R. Triboulet, Y. Marfaing: J. Cryst. Growth **51**, 89 (1981)
- 12.78 R. Triboulet, K. Pham Van, G. Didier: J. Cryst. Growth **101**, 216 (1990)
- 12.79 V. A. Kuznetsov: Prog. Cryst. Growth Ch. **21**, 163 (1990)
- 12.80 J. Mimila, R. Triboulet: Mater. Lett. **24**, 221 (1995)
- 12.81 W. W. Piper, S. J. Polich: J. Appl. Phys. **32**, 1278 (1961)
- 12.82 G. J. Russell, J. Woods: J. Cryst. Growth **46**, 323 (1979)
- 12.83 P. Blanconnier, P. Henoc: J. Cryst. Growth **17**, 218 (1972)
- 12.84 K. Durose, A. Turnbull, P. D. Brown: Mater. Sci. Eng. B **16**, 96 (1993)
- 12.85 M. R. Lorenz: *Physics and Chemistry of II–VI Compounds*, ed. by M. Aven, J. S. Prener (North-Holland, Amsterdam 1967) Chap. 2
- 12.86 K. Zanio: Semicond. Semimet. **13** (1978)
- 12.87 P. Capper, A. Brinkman: *Properties of Narrow Gap Cadmium-Based Compounds*, EMIS Datarev. Ser. **10**, ed. by P. Capper (IEE, London 1994) p. 369
- 12.88 A. W. Brinkman: *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman & Hall, London 1997)
- 12.89 R. Triboulet, J. O. Nday A. El Mokri et al.: J. Phys. IV **5**, C3–141 (1995)
- 12.90 R. Triboulet, A. Tromson-Carli, D. Lorans, T. Nguyen Duy: J. Electron. Mater. **22**, 827 (1993)
- 12.91 J. B. Mullin, C. A. Jones, B. W. Straughan, A. Royle: J. Cryst. Growth **59**, 135 (1982)
- 12.92 H. M. Hobgood, B. W. Swanson, R. N. Thomas: J. Cryst. Growth **85**, 510 (1987)
- 12.93 R. Triboulet, Y. Marfaing: J. Electrochem. Soc. **120**, 1260 (1973)
- 12.94 T. Sasaki, Y. Mori, M. Yoshimura: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 12.95 R. Hirano, H. Kurita: *Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials*, ed. by P. Capper (Wiley, Chichester 2005)
- 12.96 K. Zanio: J. Electron. Mater. **3**, 327 (1974)
- 12.97 C. P. Khatkhat, F. Schmid: Proc. SPIE **1106**, 47 (1989)
- 12.98 W. M. Chang, W. R. Wilcox, L. Regel: Mater. Sci. Eng. B **16**, 23 (1993)
- 12.99 N. R. Kyle: J. Electrochem. Soc. **118**, 1790 (1971)
- 12.100 J. C. Tranchart, B. Latorre, C. Foucher, Y. LeGouge: J. Cryst. Growth **72**, 468 (1985)
- 12.101 Y.-C. Lu, J.-J. Shiau, R. S. Fiegelson, R. K. Route: J. Cryst. Growth **102**, 807 (1990)
- 12.102 J. P. Tower, S. B. Tobin, M. Kestigian: J. Electron. Mater. **24**, 497 (1995)
- 12.103 S. Sen, S. M. Johnson, J. A. Kiele: Mater. Res. Soc. Symp. Proc. **161**, 3 (1990)
- 12.104 J. F. Butler, F. P. Doty, B. Apotovsky: Mater. Sci. Eng. B **16**, 291 (1993)
- 12.105 P. Capper, J. E. Harris, E. O'Keefe, C. L. Jones, C. K. Ard, P. Mackett, D. T. Dutton: Mater. Sci. Eng. B **16**, 29 (1993)
- 12.106 C. Szeles, S. E. Cameron, S. A. Soldner, J.-O. Nday, M. D. Reed: J. Electron. Mater. **33/6**, 742 (2004)
- 12.107 A. El Mokri, R. Triboulet, A. Lusson: J. Cryst. Growth **138**, 168 (1995)
- 12.108 R. U. Bloedner, M. Presia, P. Gille: Adv. Mater. Opt. Electron. **3**, 233 (1994)
- 12.109 R. Schoenholz, R. Dian, R. Nitsche: J. Cryst. Growth **72**, 72 (1985)
- 12.110 W. F. H. Micklethwaite: Semicond. Semimet. **18**, 3 (1981)
- 12.111 P. W. Kruse: Semicond. Semimet. **18**, 1 (1981)
- 12.112 H. Maier: *N.A.T.O. Advanced Research Workshop on the Future of Small-Gap II–VI Semiconductors* (Liege, Belgium 1988)
- 12.113 P. Capper: Prog. Cryst. Growth Ch. **19**, 259 (1989)
- 12.114 F. R. Szofran, S. L. Lehoczky: J. Cryst. Growth **70**, 349 (1984)
- 12.115 P. Capper, J. J. G. Gosney: U.K. Patent 8115911 (1981)
- 12.116 P. Capper, C. Maxey, C. Butler, M. Grist, J. Price: J. Mater. Sci. Mater. El. **15**, 721 (2004)
- 12.117 Y. Nguyen Duy, A. Durand, J. L. Lyot: Mater. Res. Soc. Symp. Proc. **90**, 81 (1987)
- 12.118 A. Durand, J. L. Dessus, T. Nguyen Duy, J. F. Barbot: Proc. SPIE **659**, 131 (1986)
- 12.119 P. Gille, F. M. Kiessling, M. Burkert: J. Cryst. Growth **114**, 77 (1991)
- 12.120 P. Gille, M. Pesia, R. U. Bloedner, N. Puhlman: J. Cryst. Growth **130**, 188 (1993)
- 12.121 M. Royer, B. R. Jean, A. R. Durand, R. Triboulet: French Patent 8804370 (1988)
- 12.122 R. U. Bloedner, P. Gille: J. Cryst. Growth **130**, 181 (1993)
- 12.123 A. Rogalski: *New Ternary Alloy Systems for Infrared Detectors* (SPIE, Bellingham 1994)
- 12.124 R. Triboulet: Semicond. Sci. Technol. **5**, 1073 (1990)
- 12.125 R. Korenstein, R. J. Olson Jr., D. Lee: J. Electron. Mater. **24**, 511 (1995)
- 12.126 B. Pellicciari, F. Dierre, D. Brellier, B. Schaub: J. Cryst. Growth **275**, 99 (2005)
- 12.127 A. Pajczkowska: Prog. Cryst. Growth Ch. **1**, 289 (1978)
- 12.128 W. Girit, J. K. Furdyna: Semicond. Semimet. **25**, 1 (1988)
- 12.129 M. C. C. Custodio, A. C. Hernandez: J. Cryst. Growth **205**, 523 (1999)
- 12.130 T. Fukuda, V. I. Chani, K. Shimamura: *Recent Developments of Bulk Crystal Growth 1998*, ed. by M. Isshiki (Research Signpost, Trivandrum, India 1998) p. 191
- 12.131 T. Fukuda, V. I. Chani, K. Shimamura: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)

- 12.132 S. Uda, S.Q. Wang, N. Konishi, H. Inaba, J. Harada: J. Cryst. Growth **237/239**, 707 (2002)
- 12.133 F. Iwasaki, H. Iwasaki: J. Cryst. Growth **237/239**, 820 (2002)
- 12.134 V. S. Balitsky: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.135 M. Korzhik: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.136 P. J. Li, Z. W. Yin, D. S. Yan: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.137 Kh. S. Bagdasarov, E. V. Zharikov: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.138 L. Lytvynov: Paper given at the 2nd International School on Crystal Growth and Technology, Zao, Japan (2000)
- 12.139 M. I. Moussatov, E. V. Zharikov: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.140 F. Schmid, Ch. P. Khattak: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.141 A. V. Gektin, B. G. Zaslavsky: Paper given at 1st International School on Crystal Growth and Technology, Beatenberg, Switzerland (1998)
- 12.142 A. V. Gektin: Paper given at the 2nd International School on Crystal Growth and Technology, Zao, Japan (2000)
- 12.143 N. Zaitseva, L. Carman, I. Smolsky: J. Cryst. Growth **241**, 363 (2002)

13. Single-Crystal Silicon: Growth and Properties

It is clear that silicon, which has been the dominant material in the semiconductor industry for some time, will carry us into the coming ultra-large-scale integration (ULSI) and system-on-a-chip (SOC) eras, even though silicon is not the optimum choice for every electronic device. Semiconductor devices and circuits are fabricated through many mechanical, chemical, physical, and thermal processes. The preparation of silicon single-crystal substrates with mechanically and chemically polished surfaces is the first step in the long and complex device fabrication process. In this chapter, the approaches currently used to prepare silicon materials (from raw materials to single-crystalline silicon) are discussed.

13.1	Overview	256
13.2	Starting Materials	257
13.2.1	Metallurgical-Grade Silicon	257
13.2.2	Polycrystalline Silicon	257
13.3	Single-Crystal Growth	258
13.3.1	Floating-Zone Method	259
13.3.2	Czochralski Method	261
13.3.3	Impurities in Czochralski Silicon ..	264
13.4	New Crystal Growth Methods	266
13.4.1	Czochralski Growth with an Applied Magnetic Field (MCZ)	266
13.4.2	Continuous Czochralski Method (CCZ)	267
13.4.3	Neckingless Growth Method	267
	References	268

Silicon, which has been and will continue to be the dominant material in the semiconductor industry for some time to come [13.1], will carry us into the ultra-large-scale integration (ULSI) era and the system-on-a-chip (SOC) era.

As electronic devices have become more advanced, device performance has become more sensitive to the quality and the properties of the materials used to construct them.

Germanium (Ge) was originally utilized as a semiconductor material for solid state electronic devices. However, the narrow bandgap (0.66 eV) of Ge limits the operation of germanium-based devices to temperatures of approximately 90 °C because of the considerable leakage currents observed at higher temperatures. The wider bandgap of silicon (1.12 eV), on the other hand, results in electronic devices that are capable of operating at up to $\approx 200^\circ\text{C}$. However, there is a more serious problem than the narrow bandgap: germanium does not readily provide a stable passivation layer on the surface. For example, germanium dioxide (GeO_2) is water-soluble and dissociates at approximately 800 °C. Silicon, in contrast to germanium, readily accommo-

dates surface passivation by forming silicon dioxide (SiO_2), which provides a high degree of protection to the underlying device. This stable SiO_2 layer results in a decisive advantage for silicon over germanium as the basic semiconductor material used for electronic device fabrication. This advantage has lead to a number of new technologies, including processes for diffusion doping and defining intricate patterns. Other advantages of silicon are that it is completely non-toxic, and that silica (SiO_2), the raw material from which silicon is obtained, comprises approximately 60% of the mineral content of the Earth's crust. This implies that the raw material from which silicon is obtained is available in plentiful supply to the IC industry. Moreover, electronic-grade silicon can be obtained at less than one-tenth the cost of germanium. All of these advantages have caused silicon to almost completely replace germanium in the semiconductor industry.

Although silicon is not the optimum choice for every electronic device, its advantages mean that it will almost certainly dominate the semiconductor industry for some time yet.

13.1 Overview

Very fruitful interactions have occurred between the users and manufacturers of semiconductor material since the invention of the point-contact transistor in 1947, when the necessity for “perfect and pure” crystals was recognized. The competition was often such that the crystal quality demanded by new devices could only be met by controlling crystal growth using electronic equipment built with these new devices. Since dislocation-free silicon crystals were grown as early as the 1960s using the *Dash technique* [13.3], semiconductor material research and developmental efforts have concentrated on material purity, production yields, and problems related to device manufacture.

Semiconductor devices and circuits are fabricated using a wide variety of mechanical, chemical, physical and thermal processes. A flow diagram for typical semiconductor silicon preparation processes is shown

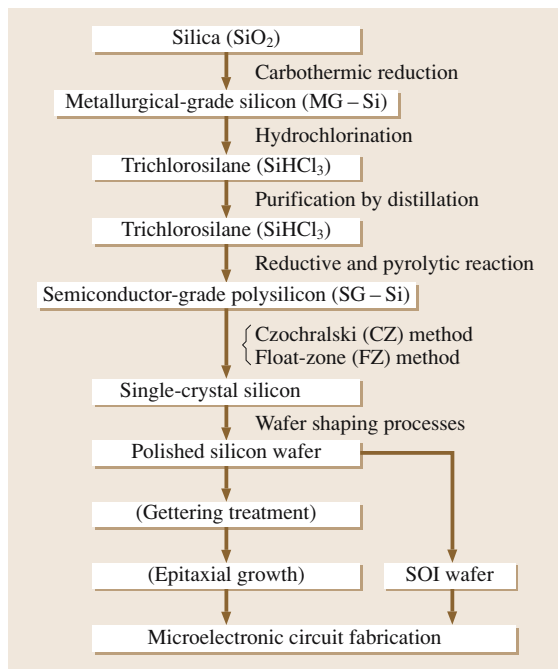


Fig. 13.1 Flow diagram for typical semiconductor silicon preparation processes. (After Shimura [13.1])

in Fig. 13.1. The preparation of silicon single-crystal substrates with mechanically and chemically polished surfaces is the first step in the long and complex process of device fabrication.

As noted above, silicon is the second most abundant element on Earth; more than 90% of the Earth’s crust is composed of silica and silicates. Given this boundless supply of raw material, the problem is then to transform silicon into the usable state required by the semiconductor technology. The first and main requirement is that the silicon used for electronic devices must be extremely pure, since very small amounts of some impurities have a strong influence on the electronic characteristics of silicon, and therefore the performance of the electronic device. The second requirement is for large-diameter crystals, since the chip yield per wafer increases substantially with larger diameters, as shown in Fig. 13.2 for the case of DRAM [13.2], one of the most common electronic devices. Besides the purity and the diameter, the cost of production and the specifications of the material, including the grown-in defect density and the resistive homogeneity, must meet current industrial demands.

In this chapter, current approaches to the preparation of silicon – converting the raw material into single-crystalline silicon (see Fig. 13.1) – are discussed.

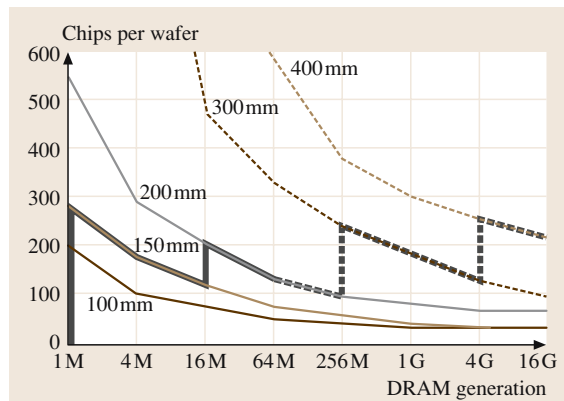


Fig. 13.2 Chips per wafer as a function of DRAM generation. (After Takada et al. [13.2])

13.2 Starting Materials

13.2.1 Metallurgical-Grade Silicon

The starting material for high-purity silicon single crystals is silica (SiO_2). The first step in silicon manufacture is the melting and reduction of silica. This is accomplished by mixing silica and carbon in the form of coal, coke or wood chips and heating the mixture to high temperatures in a submerged electrode arc furnace. This carbothermic reduction of silica produces fused silicon:



A complex series of reactions actually occur in the furnace at temperatures ranging from 1500 to 2000 °C. The lumps of silicon obtained from this process are called metallurgical-grade silicon (MG-Si), and its purity is about 98–99%.

13.2.2 Polycrystalline Silicon

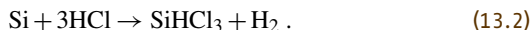
Intermediate Chemical Compounds

The next step is to purify MG-Si to the level of semiconductor-grade silicon (SG-Si), which is used as the starting material for single-crystalline silicon. The basic concept is that powdered MG-Si is reacted with anhydrous HCl to form various chlorosilane compounds in a fluidized-bed reactor. Then the silanes are purified by distillation and chemical vapor deposition (CVD) to form SG-polysilicon.

A number of intermediate chemical compounds have been considered, such as monosilane (SiH_4), silicon tetrachloride (SiCl_4), trichlorosilane (SiHCl_3) and dichlorosilane (SiH_2Cl_2). Among these, trichlorosilane is most commonly used for subsequent polysilicon deposition for the following reasons: (1) it can be easily formed by the reaction of anhydrous hydrogen chloride with MG-Si at reasonably low temperatures (200–400 °C); (2) it is liquid at room temperature, so purification can be accomplished using standard distillation techniques; (3) it is easy to handle and can be stored in carbon steel tanks when dry; (4) liquid trichlorosilane is easily vaporized and, when mixed with hydrogen, it can be transported in steel lines; (5) it can be reduced at atmospheric pressure in the presence of hydrogen; (6) its deposition can take place on heated silicon, eliminating the need for contact with any foreign surfaces that may contaminate the resulting silicon; and (7) it reacts at lower temperatures (1000–1200 °C) and at faster rates than silicon tetrachloride.

Hydrochlorination of Silicon

Trichlorosilane is synthesized by heating powdered MG-Si at around 300 °C in a fluidized-bed reactor. That is, MG-Si is converted into SiHCl_3 according to the following reaction:



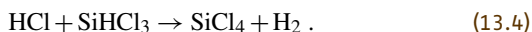
The reaction is highly exothermic and so heat must be removed to maximize the yield of trichlorosilane. While converting MG-Si into SiHCl_3 , various impurities such as Fe, Al, and B are removed by converting them into their halides (FeCl_3 , AlCl_3 , and BCl_3 , respectively), and byproducts such as SiCl_4 and H_2 are also produced.

Distillation and Decomposition of Trichlorosilane

Distillation has been widely used to purify trichlorosilane. The trichlorosilane, which has a low boiling point (31.8 °C), is fractionally distilled from the impure halides, resulting in greatly increased purity, with an electrically active impurity concentration of less than 1 ppba. The high-purity trichlorosilane is then vaporized, diluted with high-purity hydrogen, and introduced into the deposition reactor. In the reactor, thin silicon rods called slim rods supported by graphite electrodes are available for surface deposition of silicon according to the reaction



In addition this reaction, the following reaction also occurs during polysilicon deposition, resulting in the formation of silicon tetrachloride (the major byproduct of the process):



This silicon tetrachloride is used to produce high-purity quartz, for example.

Needless to say, the purity of the slim rods must be comparable to that of the deposited silicon. The slim rods are preheated to approximately 400 °C at the start of the silicon CVD process. This preheating is required in order to increase the conductivity of high-purity (high-resistance) slim rods sufficiently to allow for resistive heating. Depositing for 200–300 h at around 1100 °C results in high-purity polysilicon rods of 150–200 mm in diameter. The polysilicon rods are shaped into various forms for subsequent crystal growth processes, such as chunks for Czochralski melt growth and long cylindrical rods for float-zone growth. The process for reducing

trichlorosilane on a heated silicon rod using hydrogen was described in the late 1950s and early 1960s in a number of process patents assigned to Siemens; therefore, this process is often called the “Siemens method” [13.4].

The major disadvantages of the Siemens method are its poor silicon and chlorine conversion efficiencies, relatively small batch size, and high power consumption. The poor conversion efficiencies of silicon and chlorine are associated with the large volume of silicon tetrachloride produced as the byproduct in the CVD process. Only about 30% of the silicon provided in the CVD reaction is converted into high-purity polysilicon. Also, the cost of producing high-purity polysilicon may depend on the usefulness of the byproduct, SiCl_4 .

Monosilane Process

A polysilicon production technology based on the production and pyrolysis of monosilane was established in the late 1960s. Monosilane potentially saves energy because it deposits polysilicon at a lower temperature and produces purer polysilicon than the trichlorosilane process; however, it has hardly been used due to the lack of an economical route to monosilane and due to processing problems in the deposition step [13.5]. However, with the recent development of economical routes to high-purity silane and the successful operation of a large-scale plant, this technology has attracted the attention of the semiconductor industry, which requires higher purity silicon.

In current industrial monosilane processes, magnesium and MG-Si powder are heated to 500°C under a hydrogen atmosphere in order to synthesize magnesium silicide (Mg_2Si), which is then made to react with ammonium chloride (NH_4Cl) in liquid ammonia (NH_3) below 0°C to form monosilane (SiH_4). High-purity polysilicon is then produced via the pyrolysis of the monosilane on resistively heated polysilicon filaments at $700\text{--}800^\circ\text{C}$. In the monosilane generation process, most of the boron impurities are removed from silane via chemical reaction with NH_3 . A boron content of 0.01–0.02 ppba in polysilicon has been achieved using a monosilane process. This concentration is very low compared to that observed in polysilicon prepared from trichlorosilane. Moreover, the

resulting polysilicon is less contaminated with metals picked up through chemical transport processes because monosilane decomposition does not cause any corrosion problems.

Granular Polysilicon Deposition

A significantly different process, which uses the decomposition of monosilane in a fluidized-bed deposition reactor to produce free-flowing granular polysilicon, has been developed [13.5]. Tiny silicon seed particles are fluidized in a monosilane/hydrogen mix, and polysilicon is deposited to form free-flowing spherical particles that are an average of $700\text{ }\mu\text{m}$ in diameter with a size distribution of 100 to $1500\text{ }\mu\text{m}$. The fluidized-bed seeds were originally made by grinding SG-Si in a ball or hammer mill and leaching the product with acid, hydrogen peroxide and water. This process was time-consuming and costly, and tended to introduce undesirable impurities into the system through the metal grinders. However, in a new method, large SG-Si particles are fired at each other by a high-speed stream of gas causing them to break into particles of a suitable size for the fluidized bed. This process introduces no foreign materials and requires no leaching.

Because of the greater surface area of granular polysilicon, fluidized-bed reactors are much more efficient than traditional Siemens-type rod reactors. The quality of fluidized-bed polysilicon has been shown to be equivalent to polysilicon produced by the more conventional Siemens method. Moreover, granular polysilicon of a free-flowing form and high bulk density enables crystal growers to obtain the most from each production run. That is, in the Czochralski crystal growth process (see the following section), crucibles can be quickly and easily filled to uniform loadings which typically exceed those of randomly stacked polysilicon chunks produced by the Siemens method. If we also consider the potential of the technique to move from batch operation to continuous pulling (discussed later), we can see that free-flowing polysilicon granules could provide the advantageous route of a uniform feed into a steady-state melt. This product appears to be a revolutionary starting material of great promise for silicon crystal growth.

13.3 Single-Crystal Growth

Although various techniques have been utilized to convert polysilicon into single crystals of silicon, two techniques have dominated the production of them for

electronics because they meet the requirements of the microelectronics device industry. One is a zone-melting method commonly called the *floating-zone (FZ) method*,

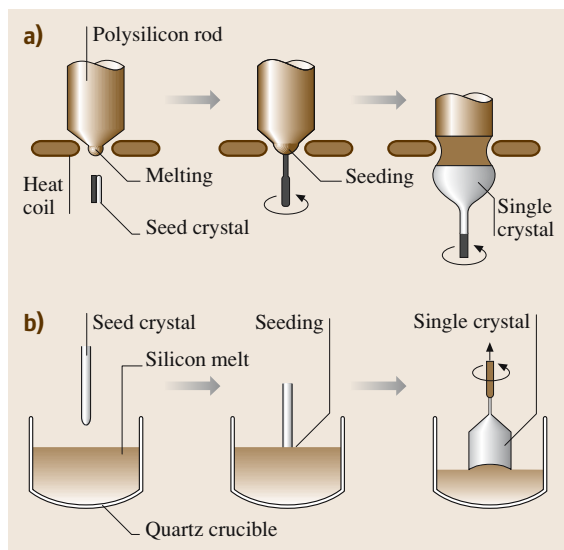


Fig. 13.3a,b Principles of single-crystal growth by (a) floating-zone method and (b) Czochralski method. (After Shimura [13.1])

and the other is a pulling method traditionally called the *Czochralski (CZ) method*, although it should actually be called the *Teal-Little method*. The principles behind these two crystal growth methods are depicted in Fig. 13.3. In the *FZ* method, a molten zone is passed through a polysilicon rod to convert it into a single-crystal ingot; in the *CZ* method, a single crystal is grown by pulling from a melt contained in a quartz crucible. In both cases, the *seed crystal* plays a very important role in obtaining a single crystal with a desired crystallographic orientation.

It is estimated that about 95% of all single-crystal silicon is produced by the *CZ* method and the rest mainly by the *FZ* method. The silicon semiconductor industry requires high purity and minimum defect concentrations in their silicon crystals to optimize device manufacturing yield and operational performance. These requirements are becoming increasingly stringent as the technology changes from LSI to VLSI/ULSI and then *SOC*. Besides the quality or perfection of silicon crystals, crystal diameter has also been steadily increasing in order to meet the demands of device manufacturers. Since microelectronic chips are produced via a “batch system,” the diameters of the silicon wafers used for device fabrication significantly affect the productivity (as shown in Fig. 13.2), and in turn the production cost.

In the following sections, we first discuss the *FZ* method and then move on to the *CZ* method. The lat-

ter will be discussed in more detail due to its extreme importance to the microelectronics industry.

13.3.1 Floating-Zone Method

General Remarks

The *FZ* method originated from zone melting, which was used to refine binary alloys [13.6] and was invented by *Theuerer* [13.7]. The reactivity of liquid silicon with the material used for the crucible led to the development of the *FZ* method [13.8], which permits the crystallization of silicon without the need for any contact with the crucible material, which is needed to be able to grow crystals of the required semiconductor purity.

Outline of the Process

In the *FZ* process, a polysilicon rod is converted into a single-crystal ingot by passing a molten zone heated by a needle-eye coil from one end of the rod to the other, as shown in Fig. 13.3a. First, the tip of the polysilicon rod is contacted and fused with a seed crystal with the desired crystal orientation. This process is called *seeding*. The seeded molten zone is passed through the polysilicon rod by simultaneously moving the single crystal seed down the rod. When the molten zone of silicon solidifies, polysilicon is converted into single-crystalline silicon with the help of the seed crystal. As the zone travels along the polysilicon rod, single-crystal silicon freezes at its end and grows as an extension of the seed crystal.

After seeding, a thin neck about 2 or 3 mm in diameter and 10–20 mm long is formed. This process is called *necking*. Growing a neck eliminates dislocations that can be introduced into newly grown single-crystal silicon during the seeding operation due to thermal shock. This necking process, called the *Dash technique* [13.3], is therefore fundamental to growing dislocation-free crystals and is used universally in both the *FZ* and the *CZ* methods. An X-ray topograph of the seed, neck and conical part of a silicon single-crystal grown by the *FZ* method is shown in Fig. 13.4. It is apparent that dislocations generated at the melt contact are completely eliminated by necking. After the conical part is formed, the main body with the full target diameter is grown. During the entire *FZ* growth process, the shape of the molten zone and the ingot diameter are determined by adjusting the power to the coil and the travel rate, both of which are under computer control. The technique most commonly used to automatically control the diameter in both the *FZ* and *CZ* methods employs an infrared sen-

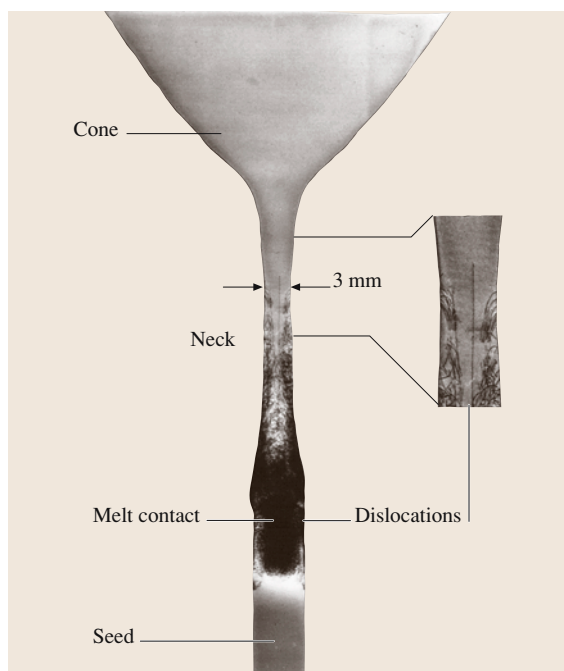


Fig. 13.4 X-ray topography of seed, neck and conical part of floating-zone silicon. (Courtesy of Dr. T. Abe)

sor focused on the meniscus. The shape of the meniscus on a growing crystal depends on its angle of contact at the three-phase boundary, the crystal diameter, and the magnitude of the surface tension. A change in meniscus angle (and therefore crystal diameter) is sensed, and the information is fed back in order to automatically adjust the growth conditions.

In contrast with CZ crystal growth, in which the seed crystal is dipped into the silicon melt and the growing crystal is pulled upward, in the FZ method the thin seed crystal sustains the growing crystal, as does the polysilicon rod from the bottom (Fig. 13.3). As a result, the rod is balanced precariously on the thin seed and neck during the entire growth process. The seed and neck can support a crystal of up to a 20 kg so long as the center of gravity of the growing crystal remains at the center of the growth system. If the center of gravity moves away from the center line, the seed will easily fracture. Hence, it was necessary to invent a crystal stabilizing and supporting technique before long and heavy FZ silicon crystals could be grown. For large crystals, it is necessary to support the growing crystal in the way shown in Fig. 13.5 [13.9], particularly in the case of recent FZ crystals with large diameters (150–200 mm), since their weights easily exceed 20 kg.

Doping

In order to obtain n- or p-type silicon single-crystals of the required resistivity, either the polysilicon or the growing crystal must be doped with the appropriate donor or acceptor impurities, respectively. For FZ silicon growth, although several doping techniques have been tried, the crystals are typically doped by blowing a dopant gas such as phosphine (PH_3) for n-type silicon or diborane (B_2H_6) for p-type silicon onto the molten zone. The dopant gas is usually diluted with a carrier gas, such as argon. The great advantage of this method is that the silicon crystal manufacturer does not need to store polysilicon sources with different resistivities.

Since the segregation (discussed in the next subsection) of elemental dopants for n-type silicon is much less than unity, FZ crystals doped by the traditional method have radial dopant gradients. Moreover, since the crystallization rate varies in the radial direction on the microscopic scale, the dopant concentrations distribute cyclically and give rise to so-called *dopant striations*, resulting in radial resistivity inhomogeneities. In order to obtain more homogeneously doped n-type silicon, neu-

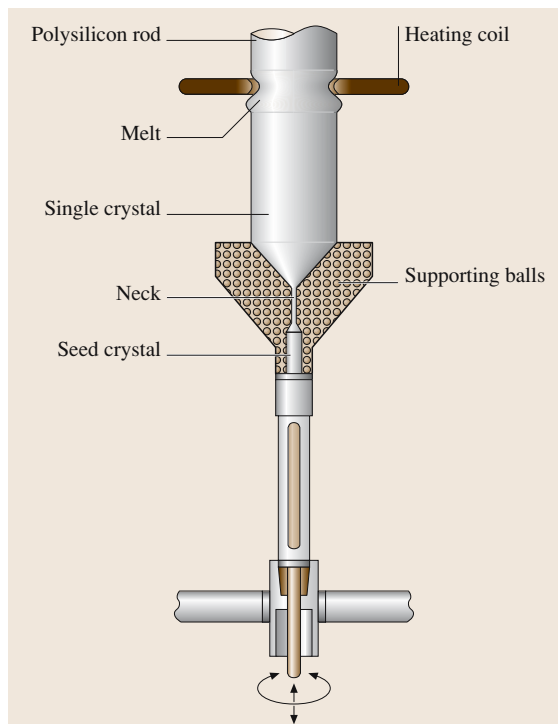
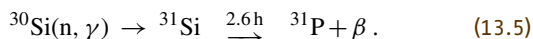


Fig. 13.5 Supporting system for floating-zone silicon crystal. (After Keller and Mühlbauer [13.9])

tron transmutation doping (NTD) has been applied to FZ silicon crystals [13.10]. This procedure involves the nuclear transmutation of silicon to phosphorus by bombarding the crystal with thermal neutrons according to the reaction



The radioactive isotope ^{31}Si is formed when ^{30}Si captures a neutron and then decays into the stable isotope ^{31}P (donor atoms), whose distribution is not dependent on crystal growth parameters. Immediately after irradiation the crystals exhibit high resistivity, which is attributed to the large number of lattice defects arising from radiation damage. The irradiated crystal, therefore, must be annealed in an inert ambient at temperatures of around 700°C in order to annihilate the defects and to restore the resistivity to that derived from the phosphorus doping. Under the NTD scheme, crystals are grown without doping and are then irradiated in a nuclear reactor with a large ratio of thermal to fast neutrons in order to enhance neutron capture and to minimize damage to the crystal lattice.

The application of NTD has been almost exclusively limited to FZ crystals because of their higher purity compared to CZ crystals. When the NTD technique was applied to CZ silicon crystals, it was found that oxygen donor formation during the annealing process after irradiation changed the resistivity from that expected, even though phosphorus donor homogeneity was achieved [13.11]. NTD has the additional shortcoming that no process is available for p-type dopants and that an excessively long period of irradiation is required for low resistivities (in the range of $1\text{--}10\ \Omega\ \text{cm}$).

Properties of FZ-Silicon Crystal

During FZ crystal growth, the molten silicon does not come into contact with any substance other than the ambient gas in the growth chamber. Therefore, an FZ silicon crystal is inherently distinguished by its higher purity compared to a CZ crystal which is grown from the melt – involving contact with a quartz crucible. This contact gives rise to high oxygen impurity concentrations of around $10^{18}\ \text{atoms}/\text{cm}^3$ in CZ crystals, while FZ silicon contains less than $10^{16}\ \text{atoms}/\text{cm}^3$. This higher purity allows FZ silicon to achieve high resistivities not obtainable using CZ silicon. Most of the FZ silicon consumed has a resistivity of between 10 and $200\ \Omega\ \text{cm}$, while CZ silicon is usually prepared to resistivities of $50\ \Omega\ \text{cm}$ or less due to the contamination from the quartz crucible. FZ silicon is therefore mainly used to fabricate semiconductor power devices that support reverse voltages

in excess of $750\text{--}1000\ \text{V}$. The high-purity crystal growth and the precision doping characteristics of NTD FZ-Si have also led to its use in infrared detectors [13.12], for example.

However, if we consider mechanical strength, it has been recognized for many years that FZ silicon, which contains fewer oxygen impurities than CZ silicon, is mechanically weaker and more vulnerable to thermal stress during device fabrication [13.13, 14]. High-temperature processing of silicon wafers during electronic device manufacturing often produces enough thermal stress to generate slip dislocations and warpage. These effects bring about yield loss due to leaky junctions, dielectric defects, and reduced lifetime, as well as reduced photolithographic yields due to the degradation of wafer flatness. Loss of geometrical planarity due to warpage can be so severe that the wafers are not processed any further. Because of this, CZ silicon wafers have been used much more widely in IC device fabrication than FZ wafers have. This difference in mechanical stability against thermal stresses is the dominant reason why CZ silicon crystals are exclusively used for the fabrication of ICs that require a large number of thermal process steps.

In order to overcome these shortcomings of FZ silicon, the growth of FZ silicon crystals with doping impurities such as oxygen [13.15] and nitrogen [13.16] has been attempted. It was found that doping FZ silicon crystals with oxygen or nitrogen at concentrations of $1\text{--}1.5 \times 10^{17}\ \text{atoms}/\text{cm}^3$ or $1.5 \times 10^{15}\ \text{atoms}/\text{cm}^3$, respectively, results in a remarkable increase in mechanical strength.

13.3.2 Czochralski Method

General Remarks

This method was named after J. Czochralski, who established a technique for determining the crystallization velocities of metals [13.17]. However, the actual pulling method that has been widely applied to single-crystal growth was developed by Teal and Little [13.18], who modified Czochralski's basic principle. They were the first to successfully grow single-crystals of germanium, 8 inches in length and 0.75 inches in diameter, in 1950. They subsequently designed another apparatus for the growth of silicon at higher temperatures. Although the basic production process for single-crystal silicon has changed little since it was pioneered by Teal and coworkers, large-diameter (up to $400\ \text{mm}$) silicon single-crystals with a high degree of perfection that meet state-of-the-art device demands have been grown

by incorporating the Dash technique and successive technological innovations into the apparatus.

Today's research and development efforts concerning silicon crystals are directed toward achieving microscopic uniformity of crystal properties such as the resistivity and the concentrations of impurities and microdefects, as well as microscopic control of them, which will be discussed elsewhere in this Handbook.

Outline of the Process

The three most important steps in CZ crystal growth are shown schematically in Fig. 13.3b. In principle, the process of CZ growth is similar to that of FZ growth: (1) melting polysilicon, (2) seeding and (3) growing. The CZ pulling procedure, however, is more complicated

than that of FZ growth and is distinguished from it by the use of a quartz crucible to contain the molten silicon. Figure 13.6 shows a schematic view of typical modern CZ crystal growth equipment. Important steps in the actual or standard CZ silicon crystal growth sequence are as follows:

1. Polysilicon chunks or grains are placed in a quartz crucible and melted at temperatures higher than the melting point of silicon (1420°C) in an inert ambient gas.
2. The melt is kept at a high temperature for a while in order to ensure complete melting and ejection of tiny bubbles, which may cause voids or negative crystal defects, from the melt.
3. A seed crystal with the desired crystal orientation is dipped into the melt until it begins to melt itself. The seed is then withdrawn from the melt so that the neck is formed by gradually reducing the diameter; this is the most delicate step. During the entire crystal growth process, inert gas (usually argon) flows downward through the pulling chamber in order to carry off reaction products such as SiO and CO.
4. By gradually increasing the crystal diameter, the conical part and shoulder are grown. The diameter is increased up to the target diameter by decreasing the pulling rate and/or the melt temperature.
5. Finally, the cylindrical part of the body with a constant diameter is grown by controlling the pulling rate and the melt temperature while compensating for the drop in the melt level as the crystal grows. The pulling rate is generally reduced toward the tail end of a growing crystal, mainly due to increasing heat radiation from the crucible wall as the melt level drops and exposes more crucible wall to the growing crystal. Near the end of the growth process, but before the crucible is completely drained of molten silicon, the crystal diameter must be gradually reduced to form an end-cone in order to minimize thermal shock, which can cause slip dislocations at the tail end. When the diameter becomes small enough, the crystal can be separated from the melt without the generation of dislocations.

Figure 13.7 shows the seed-end part of an as-grown CZ silicon crystal. Although a seed-corn, which is the transition region from the seed to the cylindrical part, is usually formed to be rather flat for economic reasons, a more tapered shape might be desirable from a crystal quality point of view. The shoulder part and its vicinity should not be used for device fabrication because this part is considered a transition region

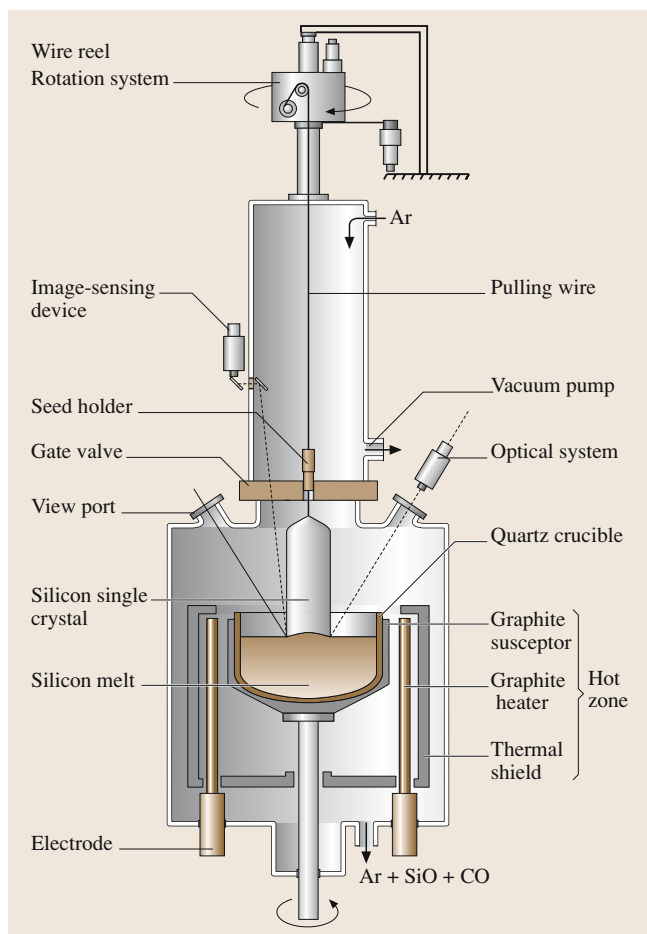


Fig. 13.6 Schematic view of typical Czochralski silicon crystal growing system. (After Shimura [13.1])

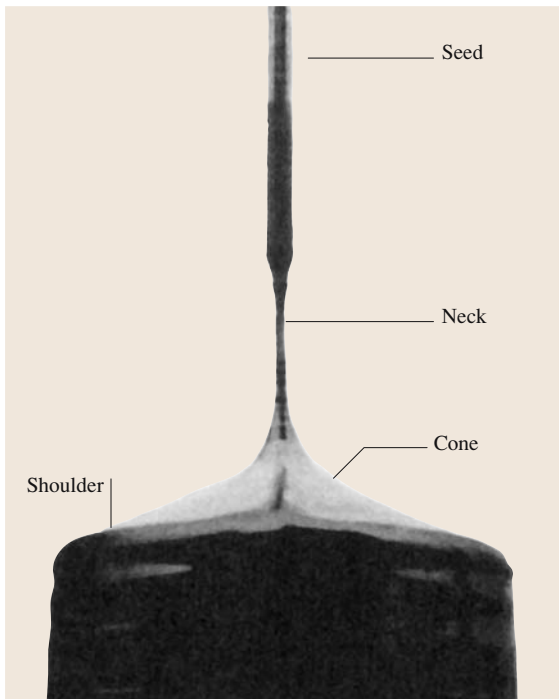


Fig. 13.7 Seed-end part of as-grown Czochralski silicon crystal

in many senses and it exhibits inhomogeneous crystal characteristics due to the abrupt change in growth conditions.

Figure 13.8 shows an extra-large as-grown CZ silicon crystal ingot 400 mm in diameter and 1800 mm in length grown by the Super Silicon Crystal Research Institute Corporation in Japan [13.2].

Influence of Spatial Location in a Grown Crystal

As Fig. 13.9 clearly depicts, each portion of a CZ crystal is grown at a different time with different growth conditions [13.19]. Thus, it is important to understand that each portion has a different set of crystal characteristics and a different thermal history due to its different position along the crystal length. For example, the seed-end portion has a longer thermal history, ranging from the melting point of 1420°C to around 400°C in a puller, while the tail-end portion has a shorter history and is cooled down rather rapidly from the melting point. Ultimately, each silicon wafer prepared from a different portion of a grown crystal could exhibit different physico-chemical characteristics depending on its location in the ingot. In fact, it has been reported that the oxygen precipitation behavior exhibits the greatest loca-

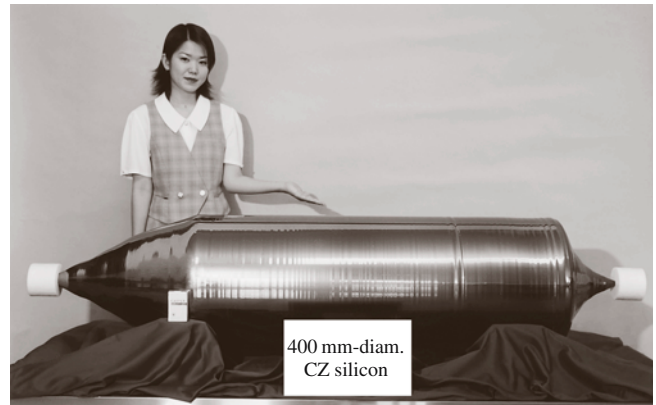


Fig. 13.8 Extra-large as-grown Czochralski silicon ingot 400 mm in diameter and 1800 mm in length. (Courtesy of Super Silicon Crystal Research Institute Corporation, Japan)

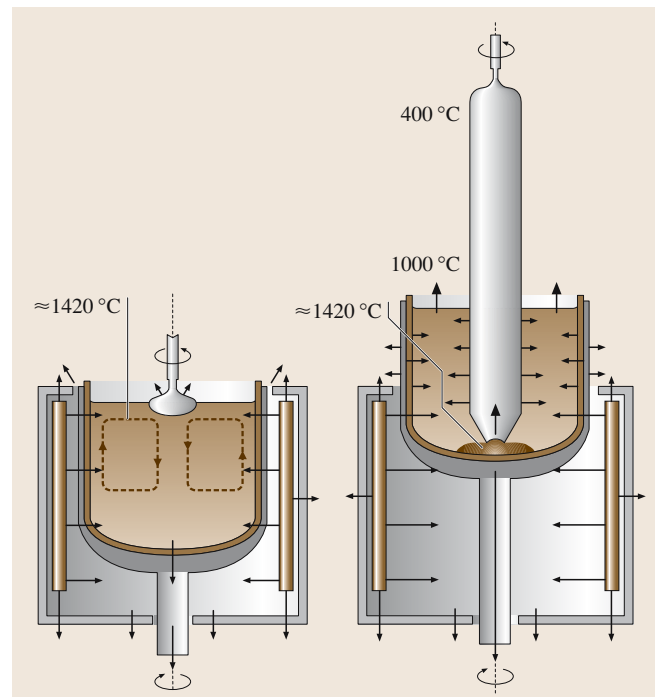


Fig. 13.9 Thermal environment during Czochralski crystal growth at initial and final stages. Arrows indicate approximate directions of heat flow. (After Zulehner and Huber [13.19])

tion dependence, which, in turn, affects the generation of bulk defects [13.20].

Also, a nonuniform distribution of both crystal defects and impurities occurs across the transverse section of a flat wafer prepared from a CZ crystal silicon melt

crystallized or solidified successively at the crystal–melt interface, which is generally curved in the CZ crystal growth process. Such inhomogeneities can be observed as *striations*, which are discussed later.

13.3.3 Impurities in Czochralski Silicon

The properties of the silicon semiconductors used in electronic devices are very sensitive to impurities. Because of this sensitivity, the electrical/electronic properties of silicon can be precisely controlled by adding a small amount of dopant. In addition to this dopant sensitivity, contamination by impurities (particularly transition metals) negatively affects the properties of silicon and results in the serious degradation of device performance. Moreover, oxygen is incorporated at levels of tens of atoms per million into CZ silicon crystals due to the reaction between the silicon melt and the quartz crucible. Regardless of how much oxygen is in the crystal, the characteristics of silicon crystals are greatly affected by the concentration and the behavior of oxygen [13.21]. In addition, carbon is also incorporated into CZ silicon crystals either from polysilicon raw materials or during the growth process, due to the graphite parts used in the CZ pulling equipment. Although the concentration of carbon in commercial CZ silicon crystals is normally less than 0.1 ppma, carbon is an impurity that greatly affects the behavior of oxygen [13.22, 23]. Also, nitrogen-doped CZ silicon crystals [13.24, 25] have recently attracted much attention due to their high microscopic crystal quality, which may meet the requirements for state-of-the-art electronic devices [13.26, 27].

Impurity Inhomogeneity

During crystallization from a melt, various impurities (including dopants) contained in the melt are incorporated into the growing crystal. The impurity concentration of the solid phase generally differs from that of the liquid phase due to a phenomenon known as *segregation*.

Segregation. The equilibrium segregation behavior associated with the solidification of multicomponent systems can be determined from the corresponding phase diagram of a binary system with a *solute* (the impurity) and a *solvent* (the host material) as components.

The ratio of the solubility of impurity A in solid silicon $[C_A]_s$ to that in liquid silicon $[C_A]_L$

$$k_0 = [C_A]_s / [C_A]_L \quad (13.6)$$

is referred to as the *equilibrium segregation coefficient*. The impurity solubility in liquid silicon is always higher than that in solid silicon; that is, $k_0 < 1$.

The equilibrium segregation coefficient k_0 is only applicable to solidification at negligibly slow growth rates. For finite or higher solidification rates, impurity atoms with $k_0 < 1$ are rejected by the advancing solid at a greater rate than they can diffuse into the melt. In the CZ crystal growth process, segregation takes place at the start of solidification at a given seed–melt interface, and the rejected impurity atoms begin to accumulate in the melt layer near the growth interface and diffuse in the direction of the bulk of the melt. In this situation, an *effective segregation coefficient* k_{eff} can be defined at any moment during CZ crystal growth, and the impurity concentration $[C]_s$ in a CZ crystal can be derived by

$$[C]_s = k_{\text{eff}} [C_0] (1 - g)^{k_{\text{eff}} - 1}, \quad (13.7)$$

where $[C_0]$ is the initial impurity concentration in the melt and g is the fraction solidified.

Consequently, it is clear that a macroscopic longitudinal variation in the impurity level, which causes a variation in resistivity due to the variation in the dopant concentration, is inherent to the CZ batch growth process; this is due to the segregation phenomenon. Moreover, the longitudinal distribution of impurities is influenced by changes in the magnitude and the nature of melt convection that occur as the melt aspect ratio is decreased during crystal growth.

Striations. In most crystal growth processes, there are transients in the parameters such as instantaneous microscopic growth rate and the diffusion boundary layer thickness which result in variations in the effective segregation coefficient k_{eff} . These variations give rise to microscopic compositional inhomogeneities in the form of *striations* parallel to the crystal–melt interface. Striations can be easily delineated with several techniques, such as preferential chemical etching and X-ray topography. Figure 13.10 shows the striations revealed by chemical etching in the shoulder part of a longitudinal cross-section of a CZ silicon crystal. The gradual change in the shape of the growth interface is also clearly observed.

Striations are physically caused by the segregation of impurities and also point defects; however, the striations are practically caused by temperature fluctuations near the crystal–melt interface, induced by unstable thermal convection in the melt and crystal rotation in an asymmetric thermal environment. In addition, mechanical vibrations due to poor pulling control mechanisms

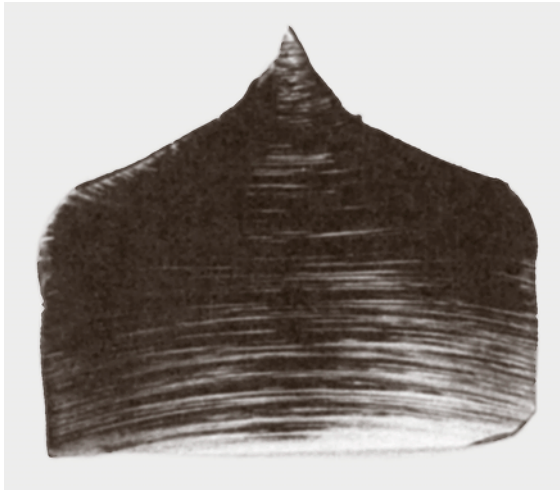


Fig. 13.10 Growth striations, revealed by chemical etching, in a shoulder of Czoehrski silicon

in the growth equipment can also cause temperature fluctuations.

Figure 13.11 schematically illustrates a CZ-grown crystal cross-section containing a curved crystal–melt interface, which results in inhomogeneities on the surface of a slice. As each planar wafer is sliced, it contains different portions of several curved striations. Different “phonograph rings”, referred to as *swirl*, can then occur in each wafer, which can be observed across the wafer using the techniques mentioned above.

Doping

In order to obtain the desired resistivity, a certain amount of dopant (either donor or acceptor atoms) is added to a silicon melt according to the resistivity–concentration relation. It is common practice to add dopants in the form of highly doped silicon particles or chunks of about $0.01\ \Omega\text{ cm}$ resistivity, which are called the dopant fixture, since the amount of pure dopant needed is unmanageably small, except for heavily doped silicon materials (n^+ or p^+ silicon).

The criteria for selecting a dopant for a semiconductor material are that it has the following properties; (1) suitable energy levels, (2) high solubility, (3) suitable or low diffusivity, and (4) low vapor pressure. A high diffusivity or high vapor pressure leads to undesirable diffusion or vaporization of dopants, which results in unstable device operation and difficulties in achieving precise resistivity control. A solubility that is too small limits the resistivity that can be obtained. In addition to

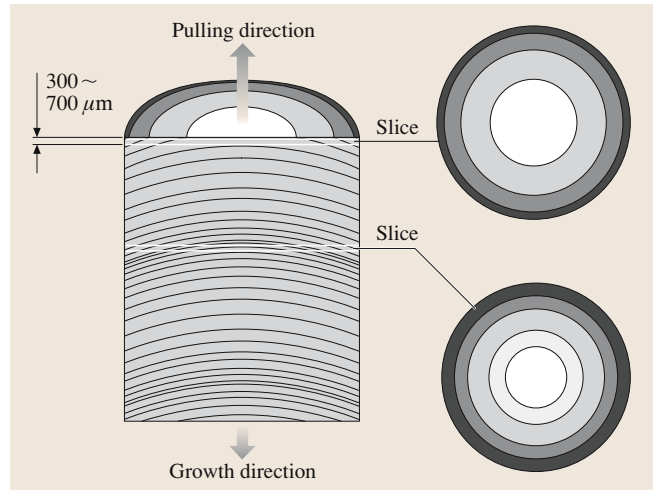


Fig. 13.11 Schematic illustration of Czoehrski crystal cross-section containing a curved crystal–melt interface and planar wafers sliced into different portions. (After Shimura [13.1])

those criteria, the chemical properties (the toxicity for example) must be considered. A further consideration from the viewpoint of crystal growth is that the dopant has a segregation coefficient that is close to unity in order to make the resistivity as uniform as possible from the seed-end to the tail-end of the CZ crystal ingot. Consequently, phosphorus (P) and boron (B) are the most commonly used donor and acceptor dopants for silicon, respectively. For n^+ silicon, in which donor atoms are heavily doped, antimony (Sb) is usually used instead of phosphorus because of its smaller diffusivity, in spite of its small segregation coefficient and high vapor pressure, which lead to large variations in concentration in both the axial and the radial directions.

Oxygen and Carbon

As shown schematically in Figs. 13.3b and 13.6, a quartz (SiO_2) crucible and graphite heating elements are used in the CZ-Si crystal growth method. The surface of the crucible that contacts the silicon melt is gradually dissolved due to the reaction



This reaction enriches the silicon melt with oxygen. Most of the oxygen atoms evaporate from the melt surface as volatile silicon mono-oxide (SiO), but some of them are incorporated into a silicon crystal through the crystal–melt interface.

However, the carbon in CZ silicon crystals originates mainly from the polycrystalline starting material. Levels

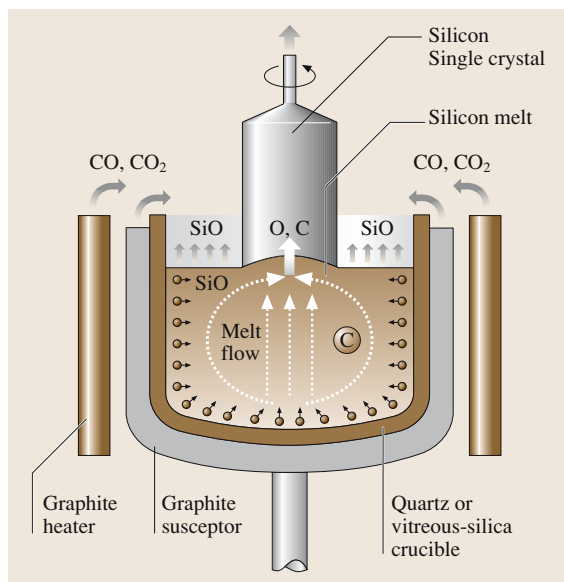


Fig. 13.12 Incorporation of oxygen and carbon into Czochralski silicon crystal. (After Shimura [13.1])

of carbon ranging from 0.1 to 1 ppma, depending on the manufacturer, are found in the polysilicon. Sources of carbon in polysilicon are assumed to be mainly carbon-containing impurities found in the trichlorosilane used in the production of polysilicon. Graphite parts in CZ pulling equipment can also contribute to carbon contamination by reacting with oxygen, which is always present during the ambient growth. The resulting products of CO and CO₂ are dissolved into the silicon melt and account for the carbon impurities in silicon crystals. Thus, oxygen and carbon are the two major nondoping impurities that are incorporated into CZ silicon crystals in the way shown schematically in Fig. 13.12. The behavior of these impurities in silicon, which affect a number of the properties of CZ silicon crystals, has been the subject of intensive study since the late 1950s [13.21].

13.4 New Crystal Growth Methods

Silicon crystals used for microelectronic device fabrication must meet a variety of requirements set by device manufacturers. In addition to the requirements for silicon wafers, the following crystallographic demands have become more common due to high-yield and high-performance microelectronic device manufacturing.

1. large diameter
2. low or controlled defect density
3. uniform and low radial resistivity gradient
4. optimum initial oxygen concentration and its precipitation.

It is clear that silicon crystal manufacturers must not only meet the above requirements but also produce those crystals economically and with high manufacturing yields. The main concerns of silicon crystal growers are the crystallographic perfection and the axial distribution of dopants in CZ silicon. In order to overcome some problems with the conventional CZ crystal growth method, several new crystal growth methods have been developed.

13.4.1 Czochralski Growth with an Applied Magnetic Field (MCZ)

The melt convection flow in the crucible strongly affects the crystal quality of CZ silicon. In particular, unfav-

orable growth striations are induced by unsteady melt convection resulting in temperature fluctuations at the growth interface. The ability of a magnetic field to inhibit thermal convection in electrically conducting fluid was first applied to the crystal growth of indium antimonide via the horizontal boat technique [13.28] and the horizontal zone-melting technique [13.29]. Through these investigations, it was confirmed that a magnetic field of sufficient strength can suppress the temperature fluctuations that accompany melt convection, and can dramatically reduce growth striations.

The effect of the magnetic field on growth striations is explained by its ability to decrease the turbulent thermal convection of a melt and in turn decrease the temperature fluctuations at the crystal–melt interface. The fluid flow damping caused by the magnetic field is due to the induced magnetomotive force when the flow is orthogonal to the magnetic flux lines, which results in an increase in the effective kinematic viscosity of the conducting melt.

Silicon crystal growth by the magnetic field applied CZ (MCZ) method was reported for the first time in 1980 [13.30]. Originally MCZ was intended for the growth of CZ silicon crystals that contain low oxygen concentrations and therefore have high resistivities with low radial variations. In other words, MCZ silicon was expected to replace the FZ silicon almost exclusively

used for power device fabrication. Since then, various magnetic field configurations, in terms of the magnetic field direction (horizontal or vertical) and the type of magnets used (normal conductive or superconductive), have been developed [13.31]. MCZ silicon produced with a wide range of desired oxygen concentrations (from low to high) has been of great interest for different device applications. The value of MCZ silicon lies in its high quality and its ability to control the oxygen concentration over a wide range, which cannot be achieved using the conventional CZ method [13.32], as well as its enhanced growth rate [13.33].

As far as the crystal quality is concerned, there is no doubt that the MCZ method provides the silicon crystals most favorable to the semiconductor device industry. The production cost of MCZ silicon may be higher than that of conventional CZ silicon because the MCZ method consumes more electrical power and requires additional equipment and operating space for the electromagnets; however, taking into account the higher growth rate of MCZ, and when superconductive magnets that need smaller space and consume less electrical power compared with conductive magnets are used, the production cost of MCZ silicon crystals may become comparable to that of conventional CZ silicon crystals. In addition, the improved crystal quality of MCZ silicon may increase production yields and lower the production cost.

13.4.2 Continuous Czochralski Method (CCZ)

Crystal production costs depend to a large extent on the cost of materials, in particular the cost of those used for quartz crucibles. In the conventional CZ process, called a *batch process*, a crystal is pulled from a single crucible charge, and the quartz crucible is used only once and is then discarded. This is because the small amount of remaining silicon cracks the crucible as it cools from a high temperature during each growth run.

One strategy for replenishing a quartz crucible with melt economically is to continuously add feed as the crystal is grown and thereby maintain the melt at a constant volume. In addition to saving crucible costs, the continuous-charging Czochralski (CCZ) method provides an ideal environment for silicon crystal growth. As already mentioned, many of the inhomogeneities in crystals grown by the conventional CZ batch process are a direct result of the unsteady kinetics arising from the change in melt volume during crystal growth. The CCZ method aims not only to reduce production costs but also to grow crystals under steady conditions. By maintain-

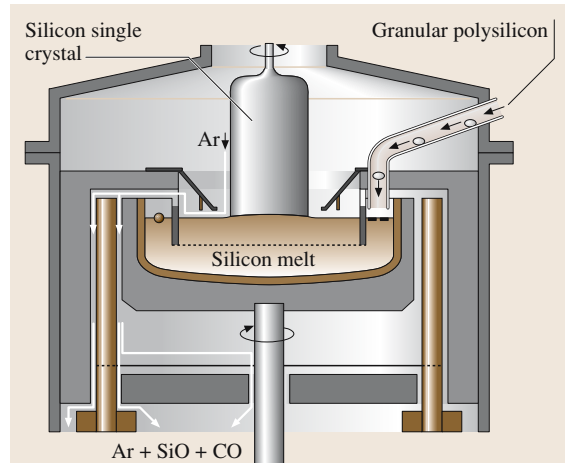


Fig. 13.13 Schematic illustration of the continuous-charging Czochralski method. (After Zulehner [13.34])

ing the melt volume at a constant level, steady thermal and melt flow conditions can be achieved (see Fig. 13.9, which shows the change in thermal environments during conventional CZ growth).

Continuous charging is commonly performed by polysilicon feeding, as shown in Fig. 13.13 [13.34]. This system consists of a hopper for storing the polysilicon raw material and a vibratory feeder that transfers the polysilicon to the crucible. In the crucible that contains the silicon melt, a quartz baffle is required to prevent the melt turbulence caused by feeding in the solid material around the growth interface. Free-flowing polysilicon granules such as those mentioned previously are obviously advantageous for the CCZ method.

The CCZ method certainly solves most of the problems related to inhomogeneities in crystal grown by the conventional CZ method. Moreover, the combination of MCZ and CCZ (the magnetic-field-applied continuous CZ (MCCZ) method) is expected to provide the ultimate crystal growth method, giving ideal silicon crystals for a wide variety of microelectronic applications [13.1]. Indeed, it has been used to grow high-quality silicon crystals intended for microelectronic devices [13.35].

However, it should be emphasized that the different thermal histories of different parts of the crystal (from the seed to the tail ends, as shown in Fig. 13.9) must be considered even when the crystal is grown by the ideal growth method. In order to homogenize the grown crystal or to obtain axial uniformity in the thermal history, some form of post-treatment, such as high-temperature annealing [13.36], is required for the crystal.

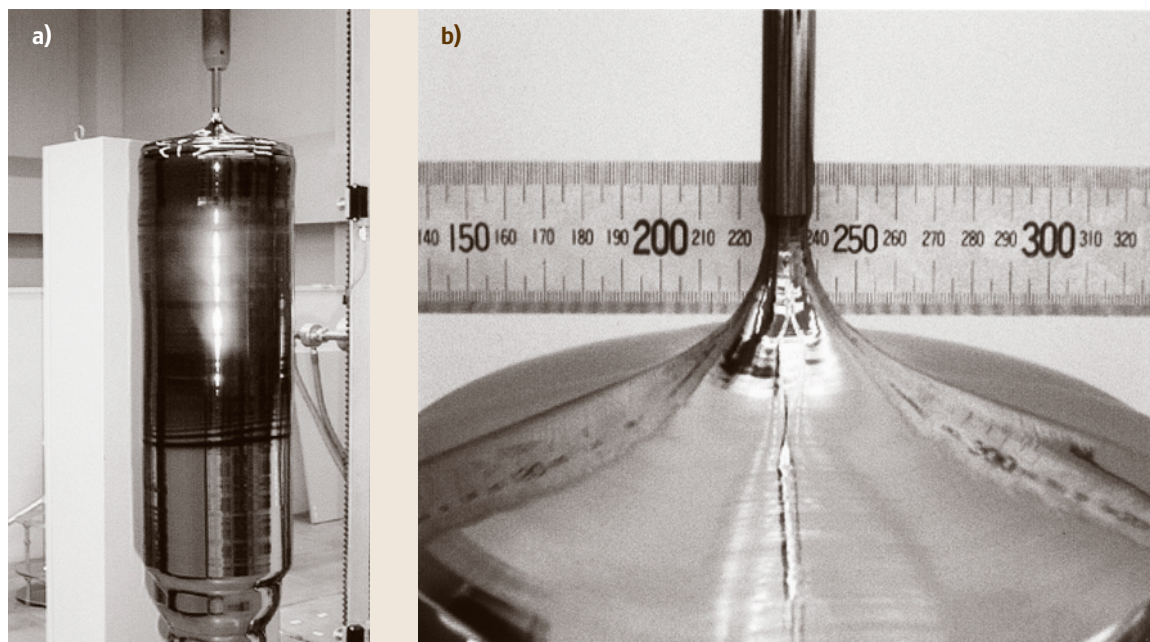


Fig. 13.14 200 mm-diameter dislocation-free Czochralski silicon crystal grown without the Dash necking process. (Courtesy of Prof. K. Hoshikawa)

13.4.3 Neckingless Growth Method

As mentioned previously, Dash's necking process (which grows a thin neck 3–5 mm in diameter, Fig. 13.7) is a critical step during CZ crystal growth because it eliminates grown-in dislocations. This technique has been the industry standard for more than 40 years. However, recent demands for large crystal diameters (> 300 mm, weighing over 300 kg) have resulted in the need for larger diameter necks that do not introduce dislocations into the growing crystal, since a thin neck 3–5 mm in diameter can not support such large crystals.

Large diameter seeds that are typically 170 mm long, with a minimum diameter of > 10 mm and

an average of 12 mm grown from silicon melt heavily doped with boron ($> 10^{19}$ atoms/cm³) have been used to grow dislocation-free 200 mm-diameter CZ silicon crystals [13.37, 38]. It is estimated that large diameter necks 12 mm in diameter can CZ support crystals as heavy as 2000 kg [13.39]. Figure 13.14a shows a 200 mm-diameter dislocation-free CZ silicon crystal grown without the Dash necking process, and Fig. 13.14b shows its enlarged seed (compare with Fig. 13.7). The mechanism by which dislocations are not incorporated into the growing crystal has been primarily attributed to hardening effect of the heavy doping of boron in the silicon.

References

- | | |
|---|--|
| <p>13.1 F. Shimura: <i>Semiconductor Silicon Crystal Technology</i> (Academic, New York 1988)</p> <p>13.2 K. Takada, H. Yamagishi, H. Minami, M. Imai: <i>Semiconductor Silicon</i> (The Electrochemical Society, Pennington 1998) p. 376</p> <p>13.3 W. C. Dash: <i>J. Appl. Phys.</i> 29, 736 (1958)</p> <p>13.4 J. R. McCormic: <i>Semiconductor Silicon</i> (The Electrochemical Society, Pennington 1986) p. 43</p> | <p>13.5 P. A. Taylor: <i>Solid State Technol.</i> 53 (1987)</p> <p>13.6 W. G. Pfann: <i>Trans. Am. Inst. Min. Metall. Eng.</i> 194, 747 (1952)</p> <p>13.7 C. H. Theuere: U.S. Patent 3,060,123,(1962)</p> <p>13.8 P. H. Keck, M. J. E. Golay: <i>Phys. Rev.</i> 89, 1297 (1953)</p> <p>13.9 W. Keller, A. Muhlbauer: <i>Floating-Zone Silicon</i> (Marcel Dekker, New York 1981)</p> |
|---|--|

- 13.10 J. M. Meese: *Neutron Transmutation Doping in Semiconductors* (Plenum, New York 1979)
- 13.11 H. M. Liaw, C. J. Varker: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1977) p. 116
- 13.12 E. L. Kern, L. S. Yaggy, J. A. Barker: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1977) p. 52
- 13.13 S. M. Hu: Appl. Phys. Lett **31**, 53 (1977)
- 13.14 K. Sumino, H. Harada, I. Yonenaga: Jpn. J. Appl. Phys. **19**, L49 (1980)
- 13.15 K. Sumino, I. Yonenaga, A. Yusa: Jpn. J. Appl. Phys. **19**, L763 (1980)
- 13.16 T. Abe, K. Kikuchi, S. Shirai: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1981) p. 54
- 13.17 J. Czochralski: Z. Phys. Chem **92**, 219 (1918)
- 13.18 G. K. Teal, J. B. Little: Phys. Rev. **78**, 647 (1950)
- 13.19 W. Zulehner, D. Hibber: *Crystals 8: Silicon, Chemical Etching* (Springer, Berlin, Heidelberg 1982) p. 1
- 13.20 H. Tsuya, F. Shimura, K. Ogawa, T. Kawamura: J. Electrochem. Soc. **129**, 374 (1982)
- 13.21 F. Shimura: *Oxygen In Silicon* (Academic, New York 1994) pp. 106, 371
- 13.22 S. Kishino, Y. Matsushita, M. Kanamori: Appl. Phys. Lett **35**, 213 (1979)
- 13.23 F. Shimura: J. Appl. Phys **59**, 3251 (1986)
- 13.24 H. D. Chiou, J. Moody, R. Sandfort, F. Shimura: *VLSI Science and Technology* (The Electrochemical Society, Pennington 1984) p. 208
- 13.25 F. Shimura, R. S. Hockett: Appl. Phys. Lett **48**, 224 (1986)
- 13.26 A. Huber, M. Kapser, J. Grabmeier, U. Lambert, W. v. Ammon, R. Pech: *Semiconductor Silicon* (The Electrochemical Society, Pennington 2002) p. 280
- 13.27 G. A. Rozgonyi: *Semiconductor Silicon* (The Electrochemical Society, Pennington 2002) p. 149
- 13.28 H. P. Utech, M. C. Flemings: J. Appl. Phys. **37**, 2021 (1966)
- 13.29 H. A. Chedzey, D. T. Hurtle: Nature **210**, 933 (1966)
- 13.30 K. Hoshi, T. Suzuki, Y. Okubo, N. Isawa: *Ext. Abstr. Electrochem. Soc. 157th Meeting* (The Electrochemical Society, Pennington 1980) p. 811
- 13.31 M. Ohwa, T. Higuchi, E. Toji, M. Watanabe, K. Homma, S. Takasu: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1986) p. 117
- 13.32 M. Futagami, K. Hoshi, N. Isawa, T. Suzuki, Y. Okubo, Y. Kato, Y. Okamoto: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1986) p. 939
- 13.33 T. Suzuki, N. Isawa, K. Hoshi, Y. Kato, Y. Okubo: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1986) p. 142
- 13.34 W. Zulehner: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1990) p. 30
- 13.35 Y. Arai, M. Kida, N. Ono, K. Abe, N. Machida, H. Futuya, K. Sahara: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1994) p. 180
- 13.36 F. Shimura: *VLSI Science and Technology* (The Electrochemical Society, Pennington 1982) p. 17
- 13.37 S. Chandrasekhar, K. M. Kim: *Semiconductor Silicon* (The Electrochemical Society, Pennington 1998) p. 411
- 13.38 K. Hoshikawa, X. Huang, T. Taishi, T. Kajigaya, T. Iino: Jpn. J. Appl. Phys **38**, L1369 (1999)
- 13.39 K. M. Kim, P. Smetana: J. Cryst. Growth **100**, 527 (1989)

14. Epitaxial Crystal Growth: Methods and Materials

The epitaxial growth of thin films of material for a wide range of applications in electronics and optoelectronics is a critical activity in many industries. The original growth technique used, in most instances, was liquid-phase epitaxy (LPE), as this was the simplest and often the cheapest route to producing device-quality layers. These days, while some production processes are still based on LPE, most research into and (increasingly) much of the production of electronic and optoelectronic devices now centers on metalorganic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE). These techniques are more versatile than LPE (although the equipment is more expensive), and they can readily produce multilayer structures with atomic-layer control, which has become more and more important in the type of nanoscale engineering used to produce device structures in as-grown multilayers. This chapter covers these three basic techniques, including some of their more common variants, and outlines the relative advantages and disadvantages of each. Some examples of growth in various important systems are also outlined for each of the three techniques.

This chapter outlines the three major epitaxial growth processes used to produce layers of material for electronic, optical and optoelectronic applications. These are liquid-phase epitaxy (LPE), metalorganic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE).

14.1	Liquid-Phase Epitaxy (LPE)	271
14.1.1	Introduction and Background	271
14.1.2	History and Status	272
14.1.3	Characteristics	272
14.1.4	Apparatus and Techniques	273
14.1.5	Group IV	275
14.1.6	Group III-V	276
14.1.7	Group II-VI	278
14.1.8	Atomically Flat Surfaces	280
14.1.9	Conclusions	280
14.2	Metalorganic Chemical Vapor Deposition (MOCVD)	280
14.2.1	Introduction and Background	280
14.2.2	Basic Reaction Kinetics	281
14.2.3	Precursors	283
14.2.4	Reactor Cells	284
14.2.5	III-V MOCVD	286
14.2.6	II-VI MOCVD	288
14.2.7	Conclusions	290
14.3	Molecular Beam Epitaxy (MBE)	290
14.3.1	Introduction and Background	290
14.3.2	Reaction Mechanisms	291
14.3.3	MBE Growth Systems	293
14.3.4	Gas Sources in MBE	295
14.3.5	Growth of III-V Materials by MBE	296
14.3.6	Conclusions	299
	References	299

We will also consider their main variants. All three techniques have advantages and disadvantages when applied to particular systems, and these will be highlighted where appropriate in the following sections.

14.1 Liquid-Phase Epitaxy (LPE)

14.1.1 Introduction and Background

Liquid-phase epitaxy (LPE) is a mature technology and has unique features that mean that it is still applicable for use in niche applications within certain device technolo-

gies. It has given way in many areas, however, to various vapor-phase epitaxy techniques, such as metalorganic vapor phase, molecular beam and atomic layer epitaxies (MOVPE, MBE, ALE), see Sects. 14.2 and 14.3. When selecting an epitaxial growth technology for a par-

ticular material system and/or device application, the choice needs to take into account the basic principles of thermodynamics, kinetics, surface energies, and so on, as well as practical issues of reproducibility, scalability, process control, instrumentation, safety and capital equipment costs. A systematic comparison of the various epitaxy techniques suggests that no single technique can best satisfy the needs of all of the material/device combinations needed in microelectronics, optoelectronics, solar cells, thermophotovoltaics, thermoelectrics, semiconductor electrochemical devices, magnetic devices and microelectromechanical systems. LPE is still a good choice for many of these application areas (M. Mauk, private communication, 2004).

14.1.2 History and Status

LPE is basically a high-temperature solution growth technique [14.1] in which a thin layer of the required material is deposited onto a suitable substrate. Homoepitaxy is defined as growth of a layer of the same composition as the substrate, whereas heteroepitaxy is the growth of a layer of markedly different composition. A suitable substrate material would have the same crystal structure as the layer, have as close a match in terms of lattice parameters as possible and be chemically compatible with the solution and the layer. Nelson [14.2] is commonly thought to have developed the first LPE systems, in this case for producing multilayer compound semiconducting structures. In the following decades a large technology base was established for III–V compound semiconductor lasers, LEDs, photodiodes and solar cells. LPE has been applied to the growth of Si, Ge, SiGe alloys, SiC, GaAs, InP, GaP, GaSb, InAs, InSb (and their ternary and quaternary alloys), GaN, ZnSe, CdHgTe, HgZnTe and HgMnTe. It has also been used to produce a diverse range of oxide/fluoride compounds, such as high-temperature superconductors, garnets, para- and ferroelectrics and for various other crystals for optics and magnetics. The early promise of garnet materials for making ‘bubble’ memories was not fully realised as standard semiconductor memory was more commercially viable. Dipping LPE is still used to make magneto-optical isolators by epitaxially growing garnet layers on gadolinium gallium garnet substrates.

It is probably true to say that most of these systems were first studied using LPE, where it was used in the demonstration, development and commercialization of many device types, including GaAs solar cells, III–V LEDs and laser diodes, GaAs-based Gunn-effect and other microwave devices and various IR detectors based

on InSb and on CdHgTe. Nevertheless, LPE does not appear in the research literature as often as, say, MOVPE, MBE and ALE in reference to work in these systems. However, it is still used extensively in industrial applications, including III–V LEDs, particularly those based on AlGaAs and GaP alloys, where it is ideally suited to the small die areas, the high luminescence efficiencies and the relatively simple device structures needed, and IR detectors based on CdHgTe.

Realistic industrial production data is difficult to obtain, but Moon [14.3] noted that the large majority of optoelectronic devices were still being grown by LPE at that time, amounting to $\approx 4000 \text{ m}^2$ per year. He also estimated that despite the loss of market share to more advanced techniques, the total demand for LPE material was still increasing at $\approx 10\%$ per year. LPE was discontinued for many applications because of its perceived limitations in regard to control of layer thickness, alloy compositions, doping, interface smoothness and difficulties in growing certain combinations of interest for heterostructure devices. LPE is normally dismissed for the production of superlattices, quantum wells, strained-layer structures and heterojunctions with large lattice mismatches of chemical dissimilarities. It also suffers from a reputation for poor reproducibility, problems with scaling up in size or throughput, and difficulties in achieving abrupt interfaces between successive layers within structures.

14.1.3 Characteristics

LPE is characterized as a near-equilibrium growth process, when compared to the various vapor-phase epitaxy techniques. Heat and mass transport, surface energies, interface kinetics and growth mechanisms are different in LPE compared to those in vapor-phase epitaxy or bulk growth techniques. These features result in both advantages and disadvantages for LPE. The former include:

- High growth rates. These are typically $0.1\text{--}10 \mu\text{m/h}$, i. e. faster than in MOVPE or MBE. This feature is useful when thick layers or “virtual substrates” are required.
- Favorable segregation of impurities into the liquid phase. This can lead to lower residual or background impurities in the epitaxial layer.
- Ability to produce very flat surfaces and excellent structural perfection (Fig. 14.1).
- Wide selection of dopants. Most solid or liquid elements can be added to a melt and incorpo-

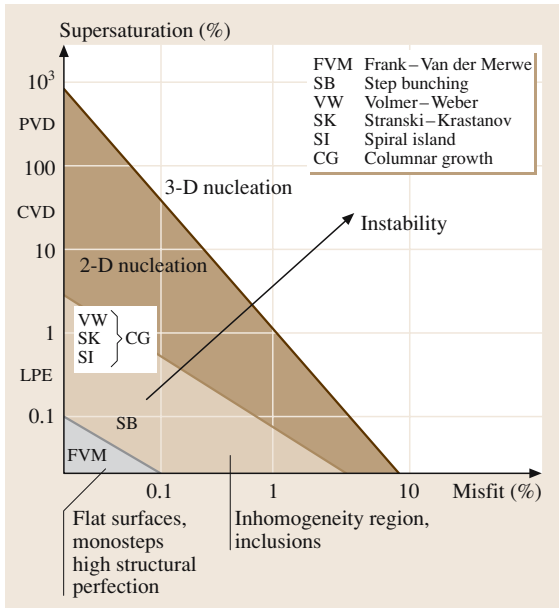


Fig. 14.1 The effects of both supersaturation (and the growth method) and misfit on the nucleation and growth regimes. Only at very low supersaturation in LPE using low-misfit substrates can really flat surfaces be expected. (After [14.4])

rated in the layer, unlike in vapor-phase growth where the development of volatile dopant precursors with suitable kinetics and sticking coefficients is a major undertaking. In this regard, there is work underway on rare-earth doping of semiconductor layers to exploit their gettering and optical properties.

- Suppression of certain types of defects. In general, LPE material has lower point defects (vacancies, interstitials, antisites) than material made by other techniques. For example, the Ga-rich conditions during GaAs LPE inhibits the formation of the As

antisite defect that is responsible for the nonradiative losses in luminescent devices.

- Once the relevant phase diagram is established, growth can be made to occur over a wide range of temperatures.
- Absence of highly toxic precursors or byproducts.
- Low capital equipment and operating costs. A research LPE kit can be constructed for under \$50 000.

The main consideration when designing an LPE process is to determine accurate phase equilibria (S–L and/or S–L–V) of the required system. Solution modeling, extrapolations of existing phase equilibria and semi-empirical predictions are usually sufficient to guide developments in new systems/applications. The near-equilibrium nature of LPE provides for several important growth modes, such as selective epitaxy (deposition through masks on a substrate) and epitaxial layer overgrowth (ELOG, where growth over a mask occurs), which are useful for defect reduction and new device structures. These new areas include work on the currently important growth of SiC and GaN for diode applications.

14.1.4 Apparatus and Techniques

The basic requirement is to bring the substrate and growth solution into contact while the epitaxial layer grows, and then to separate them cleanly at the end of the growth cycle. The three main embodiments of the LPE growth method are tipping, dipping and sliding boat, see Fig. 14.2.

Figure 14.3 shows the tipping furnace system used by Nelson [14.2]. The boat, normally graphite or silica, sits in the work tube in the center of the tilted furnace such that the substrate, held with a clamp, is held at one end of the boat with the growth solution at the other end. Once the melt has been equilibrated the temperature is

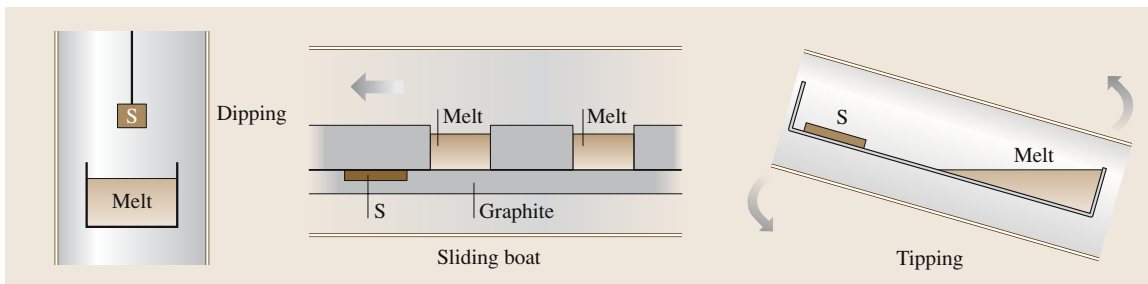


Fig. 14.2 Dipping, sliding boat and tipping LPE arrangements. (After [14.5])

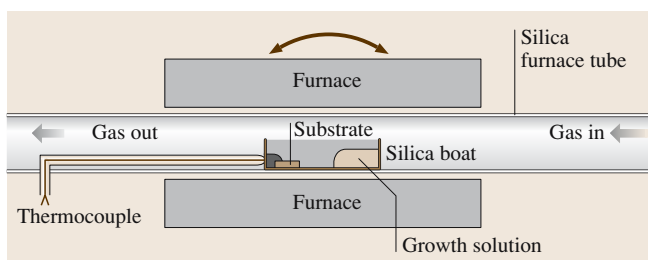


Fig. 14.3 Tipping LPE furnace. (After [14.2])

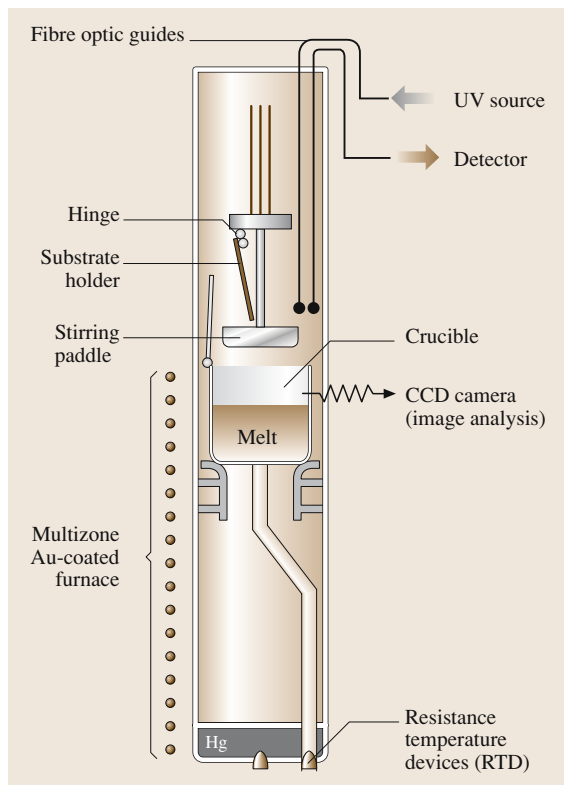


Fig. 14.4 Schematic diagram of a dipping LPE reactor showing the Te-rich melt, the mercury reservoir and positions of the sensors. (After [14.5])

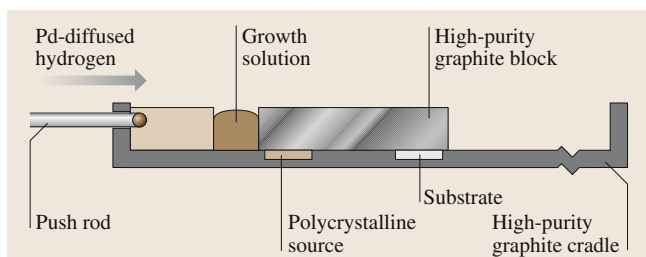


Fig. 14.5 Basic structure of graphite horizontal sliding-boat LPE. (After [14.6])

slowly reduced and the furnace is tipped to roll the solution over the substrate. After a suitable time the furnace is tipped back to the start position and the solution rolls off the grown layer. This is a relatively simple and cheap technique but has the limitations that solution removal is difficult and it is normally only suitable for single-layer growth.

Figure 14.4 shows the dipping system used for the growth of CdHgTe [14.5]. The design and operation of a system for growth from Hg-rich melts is dominated by consideration of the high vapor pressure of Hg that comprises $\geq 90\%$ of the growth solution, which led to the evolution of a vertical high-pressure furnace design with a cooled reflux region. The cylindrical melt vessel consists of a high-strength stainless steel chamber lined with quartz. Such systems are capable of containing about 10–20 kg of melt at 550 °C for several years with no degradation in melt integrity or purity. The prepared substrates are introduced into the melt through a transfer chamber or air lock. The paddle assembly can be lowered into the melt and rotated to stir the melt. In general, the high-purity melt components are introduced into the clean melt vessel at room temperature. The system is sealed, evacuated and pressurized. The temperature of the furnace is raised above the predicted melting point and held constant until all the solute dissolves. The use of large melts results in a near-constant saturation temperature from run to run and ensures excellent reproducibility of layer characteristics.

Figure 14.5 shows the basic structure of a graphite sliding-boat system, which has turned out to be the most popular and versatile of the three main methods [14.6]. The substrate sits in a recess in a slider supported by a base section. Growth solutions reside in wells in the upper section of the boat and can be repositioned over the substrate using a push-rod arrangement. One of the main drawbacks of this method is that of melt retention on the grown layer. Various means, such as empty wells, slots, lids on the solutions, and pistons to tilt the substrate have been tried with varying degrees of success. The critical design feature of the boat to aid wipe-off is to control the gap between the top of the grown layer and the underside of the top section. If this is too large, melt retention occurs, but if it is too small the layer may be scratched. Multilayer growth is easily possible using this sliding-boat method, providing melt retention is kept to a minimum. The thin melts lead to suppression of thermal or solutal convection, and hence reduce enhanced edge growth. Scale-up has also been achieved in this method with several substrates (up to 16) growing three-layer GaAlAs structures in a single run.

In all of the LPE methods, production of supersaturation in the growth solution drives the deposition of the layer on the substrate. This supersaturation can be produced by ramp cooling, step cooling, supercooled growth (a hybrid of the previous two techniques), two-phase growth, constant-temperature growth or transient growth [14.7]. The choice between these various means will depend on the details of the particular material system and the precise requirement for the material. An additional means of producing the required supersaturation is that of electroepitaxy, in which an electric current is passed through the interface to stimulate layer growth. It is now thought that this occurs via an electromigration process rather than via Peltier cooling. Benefits claimed for the technique include reduced surface ripple, a reduced number of certain microdefects and an ability to grow millimeter-thick layers of GaAlAs with uniform composition.

14.1.5 Group IV

Silicon and Silicon/Germanium

Ciszek [14.9] noted that high-quality Si layers have been grown on Si substrates at temperatures in the range 700–900 °C at a rate of 1 $\mu\text{m}/\text{min}$. The potential application was for solar cells, but because growth was on silicon, rather than a low-cost alternative, this is not considered to be a viable production process.

Alonso and Winer [14.10] grew SiGe alloys of various compositions from Si–Ge–In melts at temperatures between 640 and 900 °C. Layers were 1–5 μm thick and were used to study Raman spectra features seen in material grown by MBE. The advantage of LPE-grown material was thought to lie in the random distribution of Si and Ge atoms (no ordering is present) compared to the MBE-grown material. The authors were able to show that the Raman peaks seen in MBE-grown material were not due to ordering; rather they were due to optical phonons associated with Si–Si motion. This demonstrates the benefits that a near-equilibrium growth process can have when studying material grown by ‘nonequilibrium’ techniques.

Silicon Carbide

Dmitriev [14.8, 11] has described the production of high-quality 6H-SiC and 4H-SiC p-n junctions by LPE from Si melts. Layer thicknesses range from 0.2 to 100 μm with growth rates of 0.01–2 $\mu\text{m}/\text{min}$. Nitrogen is used as the donor impurity and aluminium, gallium and boron as acceptor impurity elements. The material showed high carrier mobility and low deep-center concentrations. Ini-

tial attempts used a technique where molten silicon ran from an upper section of the crucible to a lower section where the SiC substrates were held fixed. Dipping was also used in an attempt to grow material that was less stressed by Si melt solidification. Growth temperatures were 1500–1750 °C and layer thicknesses were 20–40 μm . Both of these techniques produced material that was successfully used to make blue LEDs. A new version of LPE, so-called container-free LPE (CFLPE) based on the electromagnetic crucible technique, was also developed, see Fig. 14.6. Liquid metal (molten Si) is suspended in a high-frequency electromagnetic field at 1000 °C and the substrates, SiC, are placed on top of the melt after heating to 1450 °C. A source of SiC is also placed at the bottom of the Si melt. Cooling of the solution was used to produce the epitaxial layer and the

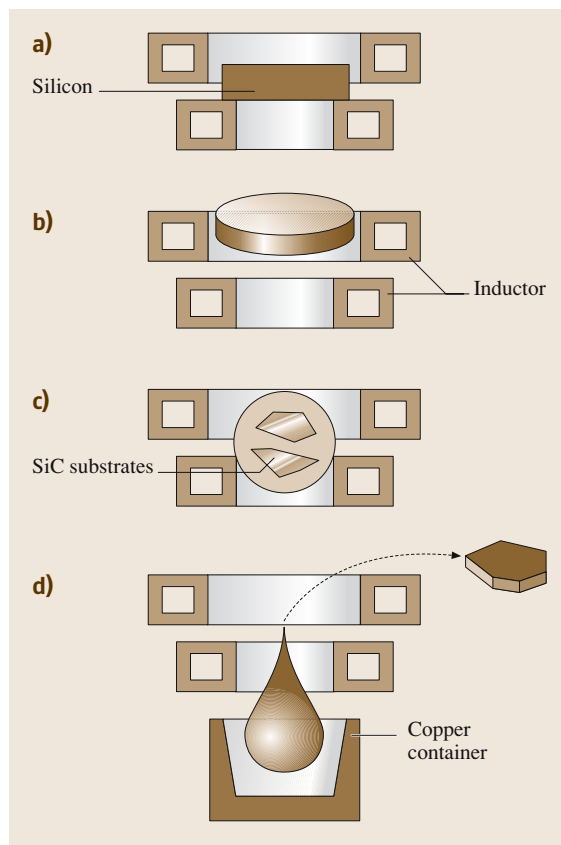


Fig. 14.6a–d Schematic of CFLPE: 1 – silicon, 2 – inductor, 3 – SiC substrate, 4 – copper container; (a) solid Si in copper inductor, (b) solid Si suspended in the inductor, (c) liquid Si with SiC crystals suspended in the electromagnetic field, (d) Si is dropped into the container. (After [14.8])

samples were then removed from the melt. The technique succeeded in producing p–n junctions by growing both layers in the same run. These formed the basis of several types of high-temperature devices (up to 500 °C), including diodes, FETs, bipolar transistors and dynistors, and optoelectronic devices such as green, blue, violet and UV LEDs. Reductions in melt temperature have been attempted by adding Sn, Ge and Ga to Si melts, with some success being reported for the latter; growth at 1100–1200 °C was obtained.

14.1.6 Group III–V

Arsenic- and Phosphorus-Based Materials

The majority of work in the area of III–V growth has been on GaAs and GaP, plus additions of As and Al. Following the earlier treatment by *Elwell* and *Scheel* [14.1], *Astles* [14.7] gave a comprehensive treatment of the LPE growth of GaAs and other III–V binaries and ternaries. He lists the advantages of LPE as: high luminescence efficiency due to the low concentration of nonradiative centers and deep levels, growth of ternary and quaternary alloys, controlled p- and n-type doping, multilayer growth with low interface recombination velocities and good reproducibility and uniformity. Disadvantages included: large areas that are required to be free of surface features (such as for photocathodes or ICs), very abrupt control of doping/composition profiles is required (as for microwave devices), accurate thickness control is required (as for microwave and quantum-well devices), and compositional grading between the substrate and the layer is inevitable. A problem associated with the use of phosphorus-containing substrates is the need to provide an overpressure source or a dummy solution to prevent phosphorus loss during the pregrowth phases.

All of the methods outlined above were attempted for the growth of GaAs and related materials. In addition, because LPE is a near-equilibrium technique that uses low supersaturation, nucleation is very sensitive to substrate lattice parameters and the growth rate is influenced by the substrate orientation. This enables localized growth in windows on the substrate surface and growth on nonplanar substrates with ribs or channels produced by preferential etching. The latter feature has been used to produce novel laser structures.

In fact, a vast array of both optoelectronic and microwave devices have been produced in LPE GaAs and related materials. The earliest were the GaAs Gunn devices and GaP/GaAsP LEDs. Later, GaAs/GaAlAs heterojunctions were produced for use in lasers, photocathodes and solar cells. Other alloy systems, such as

GaInP for blue/green LEDs, GaInSb for improved Gunn devices, and GaInAs or GaAsSb for photocathodes were also studied. Later still came growth of ternaries, such as GaInAsP (lattice-matched to InP) for heterostructure optoelectronic devices. Finally, OEICs and buried heterostructure lasers were developed to exploit the potential for selective-area growth and anisotropy of growth rate.

III–V Antimonides

Commercially available substrates for epitaxy are limited in their lattice constant spread and this imposes certain constraints in terms of lattice-matched growth and miscibility gaps. Ternary and quaternary alloy substrates with adjustable lattice parameters would open up new device applications. However, bulk-grown ternary alloys suffer from segregation and stress effects. An alternative approach is to grow very thick layers (> 50 µm) of these compounds for use as ‘virtual substrates’, *Mao* and *Krier* [14.12]. For III–V antimonides, where substrate and lattice-matching problems are acute, such thick layers are feasible by LPE due to the relatively fast growth rates (1–10 µm/min). Either gradual compositional grading or growing multilayers with abrupt but incremental compositional changes between layers can be combined with either selective removal of the substrate (to produce free-standing layers) or wafer-bonding techniques, yielding an alloy layer bonded to a surrogate substrate. The challenge for these virtual substrates is to produce lattice constants that are sufficiently different from those available using binary substrates, without introducing an excessive level of defects.

Another interesting application of antimonides is that of InSb-based quantum dots, *Krier* et al. [14.13]. The potential application here is in mid-IR lasers, LEDs and detectors. In particular, there is a market for these materials as gas detectors based on IR absorption. The principle is that of rapid slider LPE, in which a thin slit of melt is wiped across the substrate producing contact times of 0.5–200 ms. This produces low-dimensional structures such as quantum wells and quantum dots. InSb quantum dots were grown on InAs substrates at 465 °C with 10 °C supercooling and a 1 ms melt–substrate contact time. Both small (4 nm high and 20 nm in diameter) and large quantum dots (12 nm high and 60 nm in diameter) are produced. Extensions to this work included growing InSb dots on GaAs and InAsSb dots. Photoluminescence and electroluminescence in the mid-IR region ($\approx 4 \mu\text{m}$) were observed in these dots.

A Japanese group [14.14] is pioneering a technique called melt epitaxy, which can be viewed as a variant

of LPE. A sliding-boat arrangement rapidly solidifies a ternary melt into a $\approx 300\text{ }\mu\text{m}$ -thick ternary slab on a binary substrate. For example, thick InGaSb and InAsSb layers were grown onto GaAs and InAs substrates, respectively. Low background doping and high electron mobilities are achieved in material that demonstrates cut-off wavelengths in the 8–12 μm region, potentially a competitor to the more established IR detectors based on MCT (Sect. 14.1.7).

Group III Nitrides

The LPE of GaN is difficult due to the low solubility of nitrogen in molten metals at atmospheric pressure. There are reports of growth of GaN from gallium and bismuth melts, and in some instances the melt is replenished with nitrogen by introducing ammonia into the growth ambient, relying on a so-called VLS (vapor-liquid-solid) growth mechanism that essentially combines LPE with CVD (chemical vapor deposition). Another report [14.15] notes the use of Na fluxes as a solvent. *Klemenz* and *Scheel* [14.16] used a dipping mode at 900 °C with sapphire, LiGaO₂, LiAlO₂ and CVD GaN on sapphire substrates.

Other Topics

Doping with rare-earth elements (Dy, Er, Hf, Nd, Pr, Yb, Y, ...) in the AlGaAs, InGaAs, InGaAsSb and InGaAsP systems can lead to impurity gettering effects that radically reduce background doping and junction saturation currents and increase carrier mobilities and minority carrier lifetimes. Such rare-earth doping in InAsSb LEDs [14.17] increases the luminescence by 10–100 times.

There is no fundamental limit to the number of components in mixed alloy layers produced by LPE. For example, AlGaInPAs layers have been grown on GaAs by LPE [14.18]. Each additional element adds an extra degree of freedom for tailoring the properties of the layer, although more detailed phase equilibria data or models are required to determine accurate melt compositions and temperatures. However, as more constituents are added the melt becomes more dilute and more nearly approaches ideal behavior.

Traditionally, LPE melts are rich in one of the major components of the layer to be grown. However, there are certain advantages to using alternative solvents, such as bismuth, as used for GaAs. In the latter case the melt is then dilute in both arsenic and gallium and the chemical activities can be separately controlled to try to reduce point defects since the concentrations of these defects depend on the chemical potentials of the constituents.

Bismuth also has lower surface tension that provides better wetting of the substrate. Solubilities can also be changed to affect growth rates or segregation of certain elements, such as Al in AlGaAs. Other solvents that might be considered include molten salts, alloys with Hg, Cd, Sb, Se, S, Au, Ag, or even perhaps some fused oxides.

Several groups have reported success with LPE growth of several less-common semiconductors, such as InTlAsSb, InBiSb and GaMnAs [14.14]. The drive for this work is for low-bandgap material for use in detectors to rival those made in MCT (Sect. 14.1.7).

The low supersaturation of LPE makes selective modes of epitaxy feasible. A substrate can be masked (using, say, SiO₂, Si₃N₄, TiN) and patterned with openings that serve as sites for preferential nucleation. In epitaxial lateral overgrowth (ELO), the selectively seeded material overgrows the mask. This technique has been used for defect filtering, stress reduction, substrate isolation and buried mirrors and electrodes [14.14]. ELO is difficult with vapor-phase methods; aspect ratios (width to thickness of selectively grown material) are small, whereas they can be 100 in LPE. This could have potential for light-emitting diodes [14.14]. Another interesting application of selective LPE is the growth of pyramidal AlGaAs microtips for scanning near-field optical microscopy.

LPE growth of heterostructures with high lattice mismatch has also been attempted, for example of InSb on GaAs [14.14] and AlGaAs on GaP [14.14]. This can be assisted by growing a buffer layer by CVD, as in the LPE of AlGaAs on GaAs-coated (by MOCVD or MBE) silicon substrates. Defect-density reductions of ≈ 2 orders of magnitude can be achieved relative to the GaAs buffer layer grown by MOCVD or MBE.

Another variant of the basic LPE process is that of liquid-phase electroepitaxy (LPEE), where application of an electric current through the growth interface can enhance growth rates for producing thick ternary layers [14.14]. Selective LPEE on patterned, tungsten-masked GaAs substrates can produce inverted pyramid-shaped crystals that can be used to make very high efficiency LEDs [14.14].

Mauk et al. [14.19] have reported on a massive scaling up of the LPE growth of thick ($> 50\text{ }\mu\text{m}$) AlGaAs on 75 mm-diameter GaAs substrates. The method produces a two orders of magnitude improvement in areal throughput compared to conventional horizontal sliding boat systems and has applications for LEDs, thermophotovoltaic devices, solar cells and detectors. A large rectangular aluminium chamber is used instead

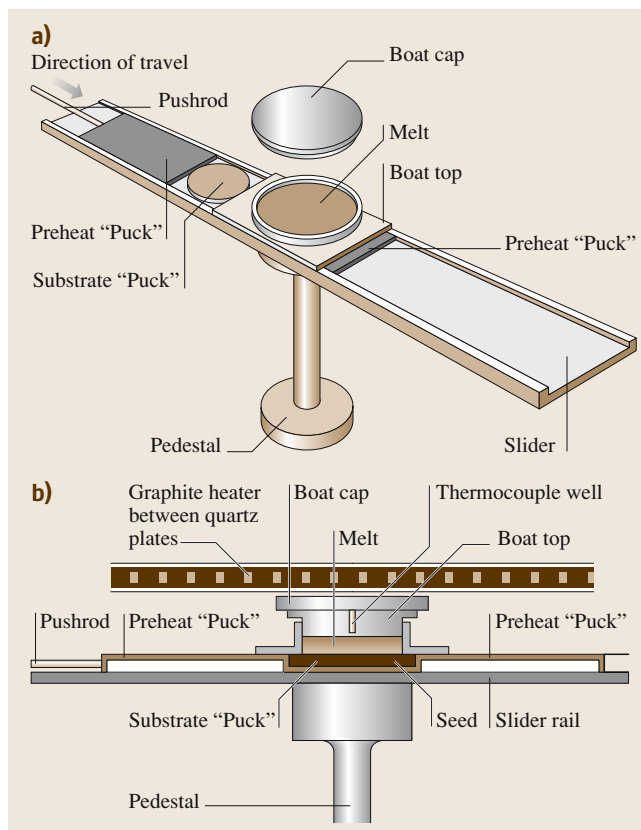


Fig. 14.7a,b Schematic of three-inch LPE apparatus, (a) perspective view, (b) side view. (After [14.19])

of a silica tube (Fig. 14.7). A modified sliding-boat arrangement is used with a top heater mounted above the boat cap, in conjunction with a heat exchanger/pedestal that acts as a cold finger to impose a vertical temperature gradient, which is the driving force for growth. These $> 50 \mu\text{m}$ -thick layers were then produced as free-standing layers bonded to glass.

14.1.7 Group II–VI

Widegap Compounds

While vapor-phase methods are normally used, LPE has been used to grow some widegap Zn-based II–VI compounds [14.20]. This work was driven by the view at that time that blue LEDs could be made economically in ZnSe. The aim was to produce p–n junctions directly by LPE via growth at $950\text{--}650^\circ\text{C}$, much lower than the bulk crystal growth, to reduce the number of defects. The problems included the high Se vapor pres-

sure, necessitating a closed-tube approach, and the need to maintain the ZnSe substrates in the upper portion of the vertically held melt during deposition. Nevertheless, $10\text{--}20 \mu\text{m}$ -thick layers were grown in 2 h. Growth at 950°C produced smoother surfaces, but the layers contained more deep levels and impurities compared with those grown at the lower temperatures. Addition of a separate Zn vapor pressure source improved the properties, showing p-type conductivity, and doping with Au, Na, and Li was also attempted. All of these produced p-type material but there was no n-type material reported. More recently, the same group [14.21] reported growth of p-type ZnSe doped with Na_2Se from which p–n junctions were fabricated after Ga diffusion from a Zn solution, to produce the n-type layer. Blue light was emitted at a wavelength of 471 nm .

Astles [14.22] has reviewed the work done on LPE of CdTe-based compounds. Most studies have been carried out from Te-rich solutions in the temperature range $500\text{--}900^\circ\text{C}$. Layers are p-type as-grown or n-type if doped with In or Al. Growth rates are typically $0.5 \mu\text{m}/^\circ\text{C}$ at 500°C . Growth from Bi-rich melts was also studied and this was found to improve melt wipe-off and surface morphology. Buffer layer growth of CdZnTe layers was used by Pellicciari et al. [14.23] as impurity barrier layers. Both CdMnTe [14.24] and HgCdMnTe [14.25] have also been grown by LPE. The latter compound was used to produce mesa diodes for room-temperature $1.3\text{--}1.8 \mu\text{m}$ applications.

Mercury Cadmium Telluride (MCT)

The situation regarding LPE of MCT was reviewed by Capper et al. [14.5]. LPE has emerged as the predominant materials growth technology for the fabrication of both first- and second-generation MCT IR focal plane arrays (FPAs). The technology has advanced to the point where material can now be routinely grown for high-performance photoconductive (PC), photovoltaic (PV) and laser detector devices covering the entire $2\text{--}18 \mu\text{m}$ spectral region. Two different technical approaches have been pursued with almost equal success: growth from Hg solutions and growth from Te solutions. One major advantage of the Hg-solution technology is its ability to produce layers of excellent surface morphology due to the ease of melt decanting. Two additional unique characteristics have now been widely recognized as essential for the fabrication of high-performance double-layer heterojunction (DLHJ) detectors by LPE: low liquidus temperature ($< 400^\circ\text{C}$), which makes a cap-layer growth step feasible, and ease of incorporating both

p-type and n-type temperature-stable impurity dopants, such as As, Sb and In, during growth.

Figure 14.4 shows the dipping system used for the growth of CdHgTe [14.5]. A typical growth procedure begins by lowering the paddle plus substrates into the melt and allowing thermal equilibrium to be reached while stirring. After reaching equilibrium, a programmed ramp reduces the melt temperature to the required level at which point the shutters are opened and the substrates are exposed to the melt. Upon completion, the paddle is withdrawn into the transfer chamber and the isolation valve is closed. Large melts allow the production of layers of up to 30 cm² with excellent compositional and thickness uniformity and allow dopant impurities to be accurately weighed for incorporation into layers and to maintain stable electrical characteristics over a long period of time. Four layers (30 cm² each) with a total area of 120 cm² can be grown in a single run [14.27]. Norton et al. [14.28] also scaled up for the growth of cap layers from Hg-rich solutions, each reactor capable of growth on four 24 cm² base layers per run.

While layers grown from Hg-rich solutions are easily doped with group VB elements with high solubility, layers grown from tellurium-rich solutions are not. Group VB dopants have low solubility and are not 100% active electrically. Group IIIB elements, indium in particular, are easily incorporated from both solutions. Indium doping from tellurium-rich melts, however, has one advantage in that the segregation coefficient is near unity.

Astles [14.22] has reviewed the experimental data of Te-rich LPE growth at 460 to 550 °C. As an exam-

ple, to compare growth parameters for Te solutions with those for Hg solutions, consider the growth of LWIR MCT ($x = 0.2$) at 500 °C from both Te and Hg solutions. The x_{Cd} for Te-rich solutions is 8.3×10^{-3} , while x_{Cd} for Hg-rich solutions is 2.6×10^{-4} . This is one of the difficulties encountered in LPE growth from Hg-rich solutions. Use of large melts, however, overcomes the Cd depletion problem. MCT epitaxial layers of the desired thickness ($> 10 \mu\text{m}$) and of uniform composition through the thickness can be grown.

A typical sliding-boat arrangement has been described by Capper et al. [14.29]. The LPE boat is made from purified, high-density, small-grain, electronic-grade graphite. The precompounded growth solution is placed in a growth well, and crushed HgTe is placed in a blind well. The HgTe acts as a buffering source for the volatile elements to help stabilize the growth solution composition and also to deliver an overpressure to the annealing well to control the metal vacancy level during annealing. These wells are connected to the annealing well with a gas channel plate that controls the movement of the volatile elements within the boat. The growth solution and HgTe overpressure source are made from high-purity elements. The control over the impurity levels in the major constituent elements is a crucial part of the control of the overall process.

At the start of the growth cycle the slider is positioned so that the substrate is under the annealing well. The loaded boat is placed in the reactor tube and the furnace is pre-heated to 520 °C and then moved over the boat. The boat heats rapidly, and after a solution melting and equilibration period, the furnace is cooled rapidly by $\approx 20^\circ\text{C}$, and then a slow cooling ramp (2 or 3 °C/h) is initiated. When the boat reaches the required growth start temperature, the slider is moved so that the substrate is positioned under the molten solution. The ramp continues until the required film thickness has been grown, after which the slider is returned to the starting position and the furnace temperature reduced rapidly to an annealing temperature. Following the anneal, the furnace is moved back to its starting position and the system is allowed to cool.

Surface morphology is controlled at two levels: microtexture and long-range variation. The microtexture is a result of misalignment of the substrate crystal plane with the growing surface. Deviations $> 0.1^\circ$ from the $\langle 111 \rangle_B$ plane lead to significant surface texture. Growth on accurately orientated substrates gives a specular surface on which atomic-scale growth features can be seen using atomic force microscopy (AFM). Figure 14.8 shows a clas-

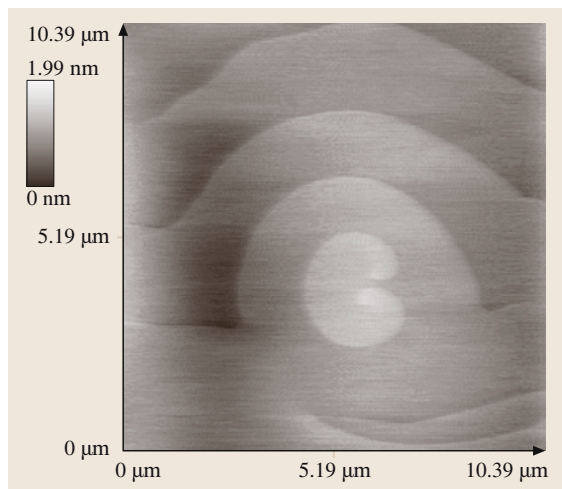


Fig. 14.8 AFM image of a Frank-Read growth spiral. (After [14.26])

sical Frank–Read site on an as-grown LPE layer surface.

The approach to forming p-on-n DLHJ structures by LPE is virtually universal. LPE from Hg-rich solution is used to grow the As- or Sb-doped p-type cap layers. The In-doped n-type base layers are grown by various Te-melt LPE techniques including tipping, sliding, and dipping. The trend appears to be in favor of the p-on-n DLHJ structures, as passivation is more controllable than that of the n-on-p structures [14.30]. A bias-selectable two-color (LWIR/MWIR) detector structure was first fabricated by growing three LPE layers from Hg-rich melts in sequence on a bulk CdZnTe substrate, Casselman et al. [14.31].

Other Narrowgap II–IV Compounds

HgZnTe was first proposed as an alternative detector material to MCT due to its superior hardness and its high energies for Hg vacancy formation and dislocation formation [14.32]. Rogalski [14.33] reviewed the LPE growth of HgZnTe and noted that Te-rich growth is favored due to the low solubility of Zn in Hg and the high Hg partial pressure. He also commented that the same factors apply to the growth of HgMnTe. Bella et al. [14.24] grew HgMnTe in a two-temperature, closed-tube tipping arrangement at 550–670 °C onto CdMnTe bulk substrates and CdMnTe LPE layers previously grown on CdTe substrates. Phase diagram data were also presented and the value of k_{Mn} was quoted as 2.5–3. Rogalski [14.33] also reviewed the status of PC and PV detectors in both HgZnTe and HgMnTe.

14.1.8 Atomically Flat Surfaces

Chernov and Scheel [14.34] have argued that far from the perceived drawback of LPE of producing rough surfaces, it may be uniquely suited to providing atomically flat, singular surfaces over distances of several micrometers. These surfaces would have applications in surface physics, catalysis and improved homogeneity of layers and superlattices of semiconductors and superconductors.

In support of this view, Fig. 14.8 shows an AFM image of a Frank–Read growth spiral on the surface of an MCT layer grown by LPE in this author's laboratory [14.26].

14.1.9 Conclusions

LPE was generally the first epitaxial technique applied to most systems of interest in micro- and optoelectronics. It is now generally a mature technology, with large fractions of several optoelectronic, IR detectors and other device types being made in LPE material, although some developments are still taking place. LPE has several advantages over the various vapor-phase epitaxial techniques, such as high growth rates, favorable impurity segregation, ability to produce flat faces, suppression of certain defects, absence of toxic materials, and low cost. There is much less emphasis on LPE in the current literature than on the vapor-phase methods, but LPE continues to seek out and develop in several niche markets where vapor-phase techniques are not suitable.

14.2 Metalorganic Chemical Vapor Deposition (MOCVD)

14.2.1 Introduction and Background

The technique of MOCVD was first introduced in the late 1960s for the deposition of compound semiconductors from the vapor phase. The pioneers of the technique, Manasevit and Simpson [14.35] were interested in a method for depositing optoelectronic semiconductors such as GaAs onto different substrates such as spinel and sapphire. The near-equilibrium techniques such as LPE and chloride VPE were not suitable for nucleation onto a surface chemically very different to the compound being deposited. These pioneers found that if they used combinations of an alkyl organometallic for the Group III element and a hydride for the Group V element, then films of GaAs could be deposited onto a variety of different surfaces. Thus, the technique of

MOCVD was born, but it wasn't until the late 1980s that MOCVD became a production technique of any significance. This success depended on painstaking work improving the impurity of the organometallic precursors and hydrides. By this time the effort was on high-quality epitaxial layers on lattice-matched substrates, in contrast with the early work. The high-quality epitaxial nature of the films was emphasized by changing the name of the growth method to metalorganic vapor phase epitaxy (MOVPE) or organometallic VPE (OMVPE). All of these variants of the name can be found in the literature and in most cases they can be used interchangeably. However, MOCVD can also include polycrystalline growth that cannot be described as epitaxy. The early niche applications of MOVPE were with GaAs photocathodes, GaAs HBT lasers and

GaInAsP lasers and detectors for 1.3 μm optical fiber communications.

The characteristics of MOCVD that have taken it from a research curiosity to production have been in the simplicity of delivery of the reactive vapors and the versatility of compositions, dopants and layer thicknesses. These basic attributes have enabled the same basic technique to be used for narrow bandgap semiconductors such as the infrared detector materials $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ and GaInSb and now for wide bandgap semiconductors such as GaN and ZnO. Indeed, the success of GaInN in the 1990s for high-brightness blue LEDs has now led to this being the most popular material produced by MOCVD. The early strength of MOCVD was its ability to grow onto different substrates but this was later abandoned in favor of the more conventional heteroepitaxy; however, the nitrides rely on heteroepitaxy onto sapphire and SiC substrates, bringing MOCVD back to its roots with the early work of Manasevit. This versatility with substrate materials presents MOCVD with the ultimate challenge of mating high-performance optoelectronic materials with silicon substrates in order to combine the best of optoelectronic and electronic performance.

This section of the chapter will cover the key elements of the MOCVD process from the physical characteristics of the precursors through reactor design to getting the right materials properties for high-performance devices.

14.2.2 Basic Reaction Kinetics

The precursors for III–V MOCVD are generally a simply alkyl for the Group III source and a hydride for the Group V source. Both have the essential properties of being volatile in a suitable carrier gas stream (usually hydrogen) and being chemically stable at ambient temperature. These precursors are normally mixed outside the reaction chamber, introduced into the reaction chamber through a suitable injector arrangement and directed onto a hot substrate. This is shown schematically in Fig. 14.9. The details of reactor design will be discussed later in this chapter. The reaction of the precursors to yield the III–V compound on the substrate can occur either in the hot vapor above the surface or on the hot surface. The stoichiometric reaction for GaAs growth is given as



This reaction has been the most widely studied of all the MOCVD reactions and was one of the original processes reported by Manasevit and Simpson [14.35]. One

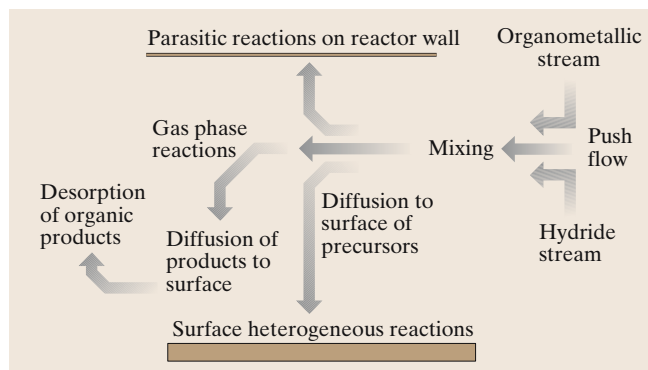


Fig. 14.9 Schematic of MOCVD process from mixing of gas streams to reaction on the substrate surface

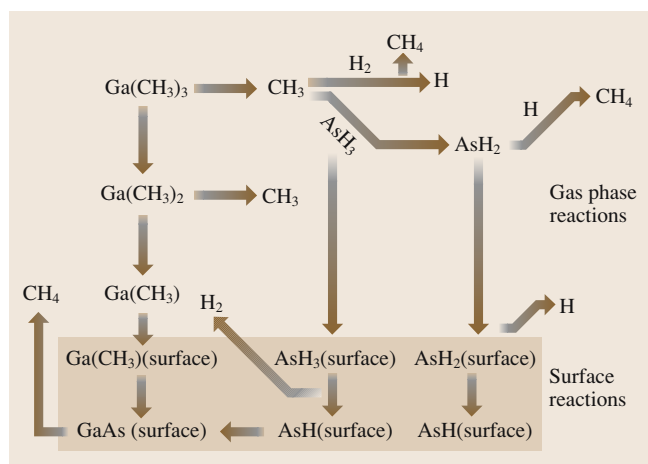


Fig. 14.10 Reaction scheme for epitaxial deposition of GaAs showing the most important vapor phase and surface reaction steps

reason that it works so well is that the hydrogen required to satisfy the $\cdot\text{CH}_3$ radical bond is supplied from the arsine hydride, and at normal growth temperatures (around 700 $^\circ\text{C}$) avoids unsatisfied carbon bonds that could lead to carbon incorporation. In reality this very simple picture covers a complexity of reaction steps that have to take place, that have been discussed in great detail by Chernov [14.36]. However, some of the important reaction steps will be described here as an introduction to the kinetics of GaAs MOCVD.

The schematic shown in Fig. 14.10 gives some of the important reaction steps that have been identified for the reaction of GaAs. This gives some insight into the complexity of the reaction kinetics and it is worth remembering that this is a relatively straightforward reaction for MOCVD. Fortunately, one does not have to

understand every step in the process before attempting to grow a layer, and this goes some way to explaining why MOCVD has developed along very empirical lines. An understanding of the reaction kinetics does, however, enable some of the problems that are associated with MOCVD to be understood, particularly when these relatively simple precursors are replaced by more complex precursors.

It can be seen from Fig. 14.10 that the reaction process is started by gas-phase homolysis of TMGa $[(\text{CH}_3)_3\text{Ga}]$ to yield dimethylgallium and methyl radicals $(\cdot\text{CH}_3)$. There are two important roles that the methyl radicals can take and this is generally important in all MOCVD processes for deposition of III–V semiconductors:

1. Methyl radicals can react with the ambient hydrogen carrier gas to yield stable methane and hydrogen radicals.
2. Methyl radicals can react with the arsine (AsH_3) to yield stable methane and AsH_2 .

Both of these steps can initiate the decomposition of arsine either through either a methyl radical or a hydrogen radical removing a hydrogen atom from the arsine; these reaction steps can be seen in Fig. 14.10. There are some important consequences of these initial reaction steps, and one you may have already spotted is that the hydrogen carrier gas is not included in the stoichiometric reaction (14.1) but can play a part in the reaction process. Another consequence is that although it is highly unlikely that the two (or more) precursors have the same

thermal decomposition characteristics, they can be used together to react at the same temperature through radical reaction mechanisms. The latter has been illustrated in many studies of different systems using ersatz reaction tube experiments, and further information on these can be found in *Stringfellow* [14.37]. For example, in the case of GaAs, the pyrolysis of TMGa will occur at 500°C , but for arsine decomposition, a temperature of around 700°C is required.

The stepwise removal of ligands from TMGa and arsine can occur in the vapor or on the surface. Figure 14.10 shows both alternatives and the dominant path will depend on both the surface temperature and vapor temperature above the substrate. A general rule is that more of the reaction process will occur on the surface at lower temperatures when the vapor reaction rate will be slower and the surface adsorption will be more efficient. It is also important to note that there are two different adsorption sites on the surface: the Ga sites that will take the anion species and the As sites that will take the cation species.

Part of the early success of GaAs and the alloy GaAlAs was due to the absence of unwanted vapor-phase reactions that could lead to the formation of polymer chains that would contaminate the layer. This was not the case with InP where the liquid alkyl source TEIn was used instead of the solid TMIn. This led to a mist in the reactor resulting from the formation of an adduct between the precursors that subsequently polymerized. This was overcome through the use of low-pressure reactors, and this has remained a feature of most production reactors today. However, the preferred precursor is TMIn and the problems of uncertain supply from a solid source has been overcome by a variety of different innovative methods. One example is to form a liquid with a stable amine, and the source then will behave in the same way as any other liquid precursor (*Frigo et al.* [14.38]).

The details of the reaction kinetics will change with substrate temperature, but as the temperature is increased a point will be reached where the rate of epitaxial growth will no longer be determined by the overall reaction rate but will be determined by the supply of precursors to the substrate. This will be reflected by a depletion of the precursor concentration immediately above the surface and a gradient in precursor concentration towards the undepleted free stream. The limitation on the rate of epitaxial growth then becomes the rate of diffusion through the depleted boundary layer to the substrate. This is called transport-limited growth and is characterized by high growth rates and only a weak dependence of growth

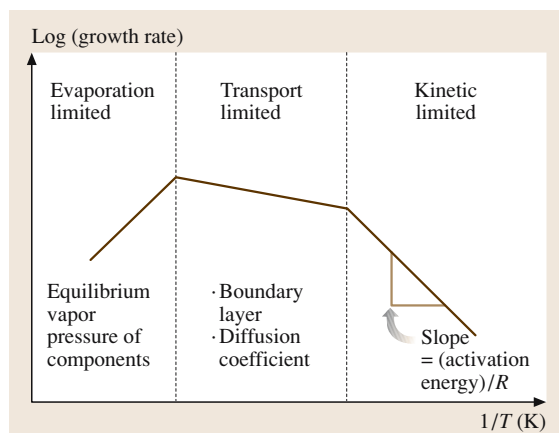


Fig. 14.11 Schematic of the three different temperature regimes for growth of a compound semiconductor by MOCVD

rate on substrate temperature. This is shown schematically in Fig. 14.11. The plot is of $\ln(\text{growth rate})$ versus $1/T$ because of the expected Arrhenius relationship in the rate constants. This really only applies to the low-temperature (kinetic) regime. Here the growth rate can be expressed as

$$\text{Rate} = A \exp(-E_a/RT), \quad (14.2)$$

where A is a constant and E_a is the activation energy. It is unlikely that E_a can be attributed to the activation energy for a single reaction step, but it is still useful for characterizing the kinetics when different precursors are being tested.

In the transport-limited regime there will be a small dependence on temperature due to the increase in diffusion rate with temperature, and this is illustrated in Fig. 14.11. Most MOCVD growth processes will take place in the transport-limited regime where it is easier to control growth rate. However, there are a number of growth processes that will occur at lower temperatures in order to control the properties such as native defect concentrations of the epitaxial films. This is generally the case with II–VI semiconductors, but can also apply to the formation of thermodynamically unstable III–V alloys.

In the high-temperature regime, the growth rate decreases with temperature, as the equilibrium vapor pressure of the constituent elements in the film will increase and give desorption rates similar to the deposition rate, leading to significant loss of material through evaporation to the gas stream.

14.2.3 Precursors

The choice of precursors is not confined to simple alkyls and hydrides but can extend to almost any volatile organometallic as a carrier for the elemental components of a film. In the case of II–VI semiconductors it is usual to use an alkyl for both the Group II and the Group VI elements. Hydrides have been used as Se and S sources but prereaction makes it difficult to control the growth process and in particular can make it difficult to incorporate dopants. The use of combined precursor sources has been extensively researched but is not in common use for epitaxial device-quality material. One reason for this is the difficulty in controlling the precursor ratio that is needed to control the stoichiometry of the material.

The important properties of precursors, and their selection, can be generalized and provides a basis for optimizing the MOCVD process. These properties can be summarized as follows:

1. Saturated vapor pressure (SVP) should be in the range of 1–10 mbar in the temperature range 0–20 °C.
2. Stable for long periods at room temperature.
3. Will react efficiently at the desired growth temperature.
4. The reaction produces stable leaving groups.
5. Avoids unwanted side reactions such as polymerization.

According to the Clausius–Clapeyron equation, the SVP of a liquid is given by an exponential relationship:

$$\text{SVP} = \exp(-\Delta G/RT), \quad (14.3)$$

where ΔG is the change in Gibbs free energy on evaporation, R is the gas constant and T the temperature of the liquid in the bubbler. This can be expressed as the heat of evaporation ΔH and the entropy for evaporation ΔS , where $\Delta G = \Delta H - T\Delta S$; this gives the familiar form of the SVP equation:

$$\begin{aligned} \text{SVP} &= \exp(-\Delta H/RT) + \Delta S/R \\ &= \exp(\Delta S/R) \exp(-\Delta H/RT). \end{aligned} \quad (14.4)$$

This is of the form:

$$\log_e(\text{SVP}) = A - B/T, \quad (14.5)$$

where A and B are constants given by $A = \Delta S/R$ and $B = \Delta H/R$. Manufacturers of the precursors will generally give the SVP data in the form of the constants A

Table 14.1 List of precursors with vapor pressure constants derived according to (14.6)

Precursor	A	B	SVP at 20 °C (mm Hg)
TMGa	8.07	1703	182
TEGa	8.08	2162	5.0
TMAI	8.22	2134	8.7
TEAI	9.0	2361	0.02
TMIn	10.52	3014	1.7
TEIn	8.94	2815	1.2
Solution TMIn	10.52	2014	1.7
DMZn	7.80	1560	300
DEZn	8.28	2109	12
DMCd	7.76	1850	28.2
DES	8.184	1907	47
DMSe	9.872	2224	
DESe	8.20	2020	
DMTe	7.97	1865	40.6
DIPTe	8.29	2309	2.6

and B in (14.5). It can also be given in the form

$$\log_{10}(\text{SVP}) = A' + B'/T. \quad (14.6)$$

To convert the constants in (14.6) to (14.5), just multiply by $\ln 10$. Some examples of the SVP constants, along with the calculated SVP at 20 °C, for a number of typical precursors are shown in Table 14.1.

14.2.4 Reactor Cells

The design of reactor cells has formed a very important part of the development of MOVPE and has been crucial in scaling laboratory processes to large-scale production. The original research reactors fell into one of two groups, either the vertical reactor or horizontal reactor. These reactor designs are shown schematically in Fig. 14.12. The substrate is placed onto a graphite susceptor that is heated by either RF coupling via a coil surrounding the reactor, a resistance heater underneath the susceptor, or lamps placed underneath the susceptor. The reactor wall can be water-cooled or gas-cooled to minimize reaction and deposition onto these surfaces. Either of these reactor cells could be operated at at-

mospheric or reduced pressure. For reduced-pressure operation the reactor cell pressure would be typically a tenth of an atmosphere but a wide range of different pressures have been successfully used. Reduced pressure will increase the gas velocity and help to overcome the effects of free convection from a hot substrate. The forced convection parameter that is often quoted is the Reynolds number and is proportional to gas velocity. A high Reynolds number will ensure streamline flow, while at low Reynolds number the buoyancy effects of the hot substrate will take over and the gas flow will be dominated by free convection (characterized by the Grashof number) and become disorganized with recirculation cells. In transport-limited growth the exact nature of the gas flow will determine the uniformity of deposition and can also affect the defect concentration in the films due to particulates and reaction products being swept back across the growing surface. It is not normal to achieve the very high flow velocities and Reynolds numbers associated with turbulent flow in an MOVPE chamber, but the disorganized flow due to free convection is often (wrongly) referred to as 'turbulent flow'.

Another reason for using high flow velocities is to overcome the effects of depletion of the precursor concentration at the downstream end of the deposition region. For transport-limited growth the growth rate is limited by the rate of diffusion from the free stream to the substrate. This region is called the boundary layer and increases in thickness going downstream from the leading edge of the susceptor, as shown in Fig. 14.13. Some horizontal reactors are designed with a tilt in the susceptor so that the free cross-sectional area decreases and hence the flow velocity increases going downstream. This helps to flatten the boundary layer and ensure better uniformity.

Maintaining a boundary layer has a cost: the high utilization of expensive precursors and gases, as the

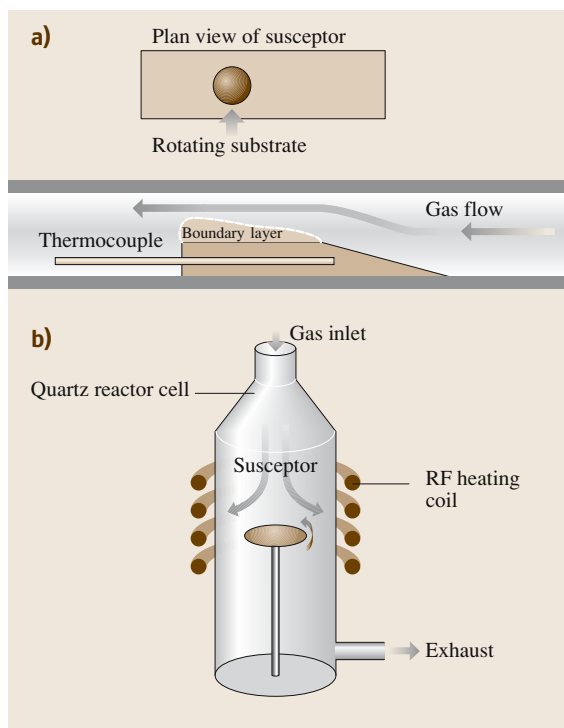


Fig. 14.12a,b Schematic of (a) horizontal and (b) vertical reactor cells

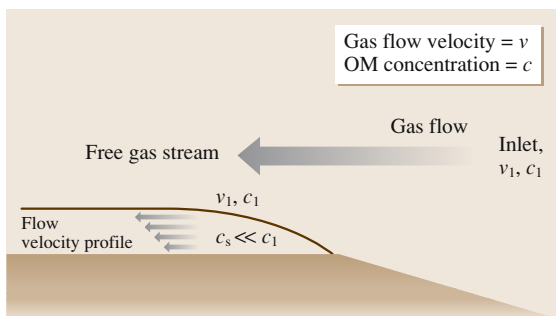


Fig. 14.13 Schematic of a boundary layer in a horizontal reactor

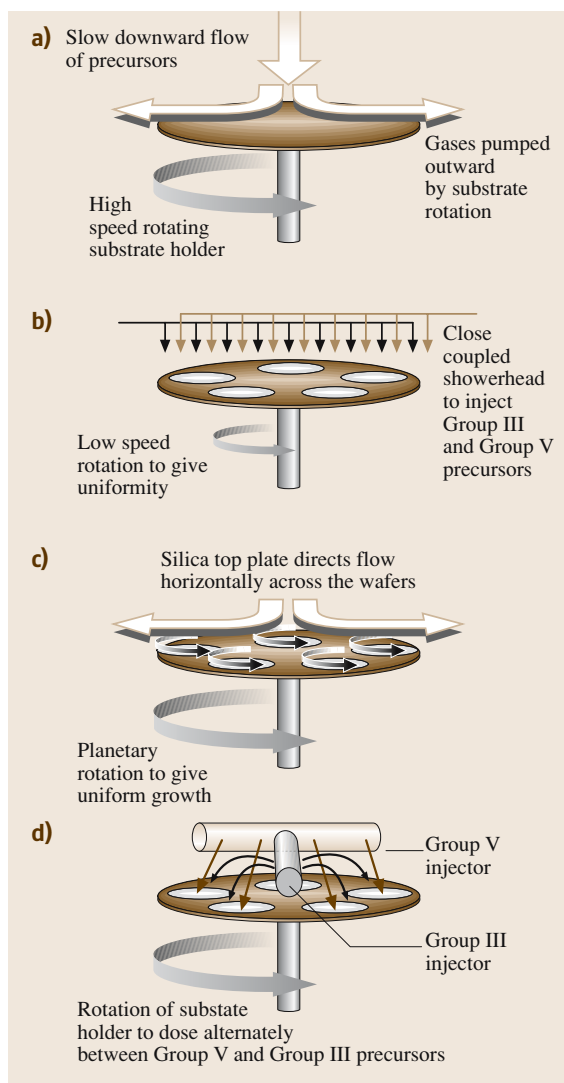


Fig. 14.14a–d Schematic of four different production reactor designs: (a) the Emcore (Veeco) Turbo-disc, (b) the Thomas Swan showerhead, (c) the Aixtron planetary, and (d) the EMF Vector flow reactor

flow throughput is typically very high and the proportion of precursors reacted in the region of the substrates is very low. This became a more serious problem when reactors were scaled to hold multiple wafers for production. These problems have been resolved with different solutions in the vertical and horizontal reactor configurations with the Emcore (now Veeco) turbo-disc reactor, the Thomas Swan showerhead reactor and the Aixtron planetary reactor. A fourth production-scale reactor de-

sign has recently been introduced by EMF, called the Titan, and works on the principle of atomic layer deposition (ALD), where the Group II and Group V gases are fed over the substrate surface separately. A common feature to all of these production reactor designs is that the substrates are rotated so that the concentrations of precursors and reactants arriving at the substrate do not have to be uniform across the surface, as a portion of the substrate will alternately experience high and low concentrations that will average out. Each of these reactor designs is shown schematically in Fig. 14.14.

The turbo-disc reactor, shown schematically in Fig. 14.14a, is a vertical reactor configuration but the boundary layer is kept to a narrow region above the susceptor by high-speed rotation that pumps the gas radially outwards due to viscous drag. The rotation speeds are up to 2000 rpm in order to create this lateral flow of the constituents above the substrate. This is continuously replenished from the slower downward gas stream, resulting in excellent uniformity of deposition across the wafers and a high utilization of the reactant gases (Tompa et al. [14.39]). The reactor pressure is typically around 100 mbar.

The showerhead reactor is another vertical reactor arrangement but it takes a different approach to overcoming free-convection currents and poor uniformity [14.40]. The precursors are introduced through a water-cooled showerhead placed just above the susceptor. The susceptor is rotated but typically at much lower speeds than for the turbo-disc reactor. The precursor distribution can be balanced across the width of the reactor to give a uniform supply of precursors.

The planetary reactor is a horizontal flow arrangement where the reactants enter at the center of rotation of the susceptor and flow outwards. This is an example of a fully developed flow where depletion of the reactants is occurring as the gases move away from the center and this will be accentuated by a decrease in the mean flow velocity as the gases move outwards [14.41]. This would normally give very poor uniformity but the planetary rotation mechanism will rotate each wafer on the platen so it will sample alternately high and low concentrations, giving uniform deposition. This approach has the advantage of high utilization of the precursors and the ability to extend the design to very large reaction chambers for multiple wafers, with the Aixtron 3000 reactor holding 95 2 inch-diameter wafers.

The fourth approach to multiple wafer deposition is the EMF Ltd vector flow epitaxy (VFE), which introduces the Group II and Group V precursors separately over a rotating susceptor platen, as shown in Fig. 14.14d.

The rotation of the platen will direct the gases across the wafers and out through separate exhausts, thus keeping the gases separate in the reactor chamber. This has the advantage of alternately dosing the surface with Group III and Group V precursors to grow the film from atomic layers, which in turn prevents prereaction between the precursors and maintains excellent film uniformity over the growth surface. This could be particularly important for compounds of nitrides and oxides where reduced pressure is normally required to avoid significant prereaction. The advantage of the **ALD** approach is that the reaction chamber can be operated at atmospheric pressure, which simplifies the operation of the system.

14.2.5 III–V MOCVD

This section will consider the range of III–V materials grown by **MOCVD** and the precursors used. Most of the III–V semiconductors can be grown from organometallics of the Group III element and hydrides of the Group V element. Exceptions to this will be noted where appropriate.

Arsenides and Phosphides

The most commonly studied alloy system is $\text{Al}_{1-x}\text{Ga}_x\text{As}$, which is used for **LEDs** and laser diodes from the near-infrared to the red part of the visible spectrum. This is a well-behaved alloy system with only a small change in lattice parameter over the entire composition range and it covers a range of bandgaps from 1.435 eV for GaAs to 2.16 eV for AlAs. One problem with this alloy is the sensitivity of aluminium to oxygen, which makes it extremely difficult to grow high-quality AlAs. Just 1 ppm of oxygen contamination will result in 10^{20} cm^{-3} incorporation of oxygen into $\text{Al}_{0.30}\text{Ga}_{0.7}\text{As}$ [14.42]. In addition to the normal **MOCVD** precautions of using ultrahigh-purity hydrogen carrier gas and ensuring that the moisture in the system is removed, the hydrides and organometallics also need to have extremely low oxygen contents. Precursor manufacturers have tended to keep to the simple alkyl precursors but to find innovative ways of reducing the alkoxide concentrations.

Alternative Group V precursors have been sought due to the high toxicity of arsine and phosphine. These hydride sources also suffer from the fact that they are stored in high-pressure cylinders and any leakage could result in the escape of large quantities of toxic gas. Alternative alkyl Group V sources have been extensively researched but only two precursors have proved to be suitable for high-quality epitaxial growth, tertiarybutyl-

arsine (**TBA**s) and tertiarybutylphosphine (**TBP**). These precursors only have one of the hydrogen ligands replaced with an alkyl substituent but they are liquid at room temperature rather than high-pressure gases. In the reactor chamber the likely reaction path is to form the hydride by a process called beta-hydrogen elimination. This entails one of the hydrogen atoms from the methyl groups satisfying the bond to As (or P) with a butene leaving group as shown below [14.43]:



This process is more likely to dominate at the normal growth temperature for transport-limited growth and it effectively yields the arsine precursor that can then react in the normal way. In the search for alternative alkyl precursors this proved to be an important factor, as the fully substituted alkyl arsenic sources tended to incorporate large concentrations of carbon, degrading the electrical properties of the film. The importance of the Group V hydride was discussed in Sect. 14.2.2 and it can be understood why **TBA** and **TBP** (for the phosphorus alloys) have proved to be good alternatives to the hydrides. However, it is fair to say that these have never been widely utilized due to much higher cost than the hydrides and poor availability.

An alternative for improved safety has been investigated more recently and relies on the same principle of reducing the toxic gas pressure in the event of a system leak. This alternative stores the hydride in a reversible adsorption system [14.44]. The adsorption system keeps the hydride at sub-atmospheric pressure and requires pumping to draw off the hydride when needed, making it inherently safer. One major advantage to this system, in addition to the inherent improvement in safety, is that the precursors and hence the precursor chemistry are unchanged in the reactor cell.

Other alloys commonly grown using **MOCVD** include $\text{In}_{0.5}\text{Ga}_{0.5}\text{P}$, which has a band gap of approximately 2 eV and is lattice-matched to GaAs. The quaternary alloy GaInAsP enables lattice-matching to InP substrates while controlling the bandgap in the 1.3 μm and 1.55 μm bands used for long-range fiber-optic telecommunications.

Antimonides

The antimonides cover an important range of bandgaps from the near-infrared to the mid-infrared bands, up to 5 or 6 μm . These compounds and alloys can be used in infrared detectors, thermophotovoltaic (**TPV**) devices and high-speed transistors. The growth of the antimonides is more complex than for the arsenides and phosphides be-

cause the hydride, stibine, is not very stable at room temperature so the use of alkyl precursors has been a more natural choice. Another factor that has influenced the growth of the antimonides is the lower thermodynamic stability and decomposition of substrates such as InSb above about 400 °C. It is also desirable to grow the films at a much lower temperature than for the arsenides and phosphides in order to keep the native defect concentration low for controlled n-type and p-type doping. However, the antimonides have an advantage in that they do not incorporate carbon as readily as in the arsenides and phosphides and there is greater flexibility over the choice of antimony precursors [14.45]. The easiest choice is to use trimethylantimony (TMSb) or to use larger alkyl groups such as triethyl and triisopropyl to reduce the reaction temperature as required. It is also possible to reduce the reaction temperature by replacing TMGa with triethylgallium (TEGa) [14.46]. This approach has been particularly advantageous when growing aluminium-containing alloys where carbon incorporation can be a problem, but is reduced using TEGa. This serves to illustrate the flexibility of MOCVD and has given more scope for the design of precursors, not envisaged in the early days of MOCVD.

Nitrides

The nitrides, mainly GaN and the alloy GaInN have brought MOCVD into prominence as a manufacturing technology with the success of high-brightness blue LEDs making large-screen full-color LED displays a reality. The precursors used for the nitrides are standard with the methyl alkyls (TMGa and TMIIn) for the Group III elements and ammonia for the nitrogen. The key technological barriers to obtaining device-quality GaN were to overcome the problems associated with heteroepitaxy onto a non-lattice-matched substrate, sapphire or SiC and to control p-type doping [14.47]. Some research has been carried out with homoepitaxy onto GaN substrates, but the very high pressures needed for bulk crystal growth of GaN will restrict the sizes of substrates available.

The heteroepitaxial problems have been overcome with a two-stage growth. High-quality GaN requires growth temperatures in excess of 1000 °C, much higher than is needed for the arsenides and phosphides. At these temperatures, nucleation onto sapphire is poor and large faceted islands grow before complete coalescence of the film occurs. This not only leads to very poor surface morphology but a high dislocation density where the islands coalesce. Two-stage growth overcomes this by growing a uniform nucleation (or buffer) layer onto the

sapphire at 600 °C. This is then heated to normal growth temperature, where a thicker GaN film is then grown. The whole nucleation process can be monitored in situ using laser reflectometry and an example is shown in Fig. 14.15 [14.48]. It can be seen that approximately 50 to 100 nm of GaN is grown as a smooth layer but during heating this changes to a rough layer. A remarkable part of the process is that during the high-temperature growth, shown in Fig. 14.15 by interference oscillations, there is a recovery in the surface morphology, resulting in smooth, device-quality layers. In some cases the GaN buffer layer is substituted with an AlN layer, but in all cases a two-stage growth process is required. The growth of device layers is achieved by controlling the band gap through growth of the alloys GaInN or GaAlN. The band gap of GaN is 3.4 eV, which is in the UV, so the color of the LED is determined by the alloy composition where increasing the In content will reduce the band gap and push the emission wavelength from blue to green. However, this is not an easy alloy to form due to the different stabilities of the GaN and InN bonds [14.49]. In fact, it cannot be grown to any useful In content at temperatures above 1000 °C, so the temperature must be reduced to around 800 °C. Even with this compromise, the different lattice parameters of GaN and InN result in poor solubility and In contents of more than 40% are not practical. In principle it is possible to prepare LEDs of any color by just changing the In content in the alloy, but in practice the high-brightness diodes can only be prepared from nitrides covering the blue to green portion of the spectrum. The growth of Al-containing alloys is

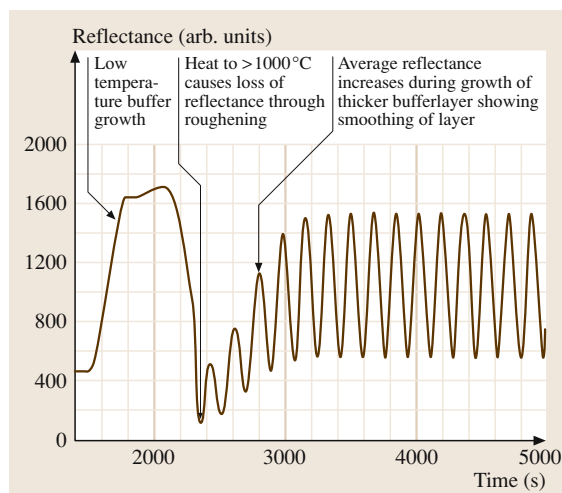


Fig. 14.15 Laser reflectometry of GaN nucleation and growth

desirable in order to achieve wider band gaps for applications such as **UV LEDs** and solar-blind detectors. The problems are similar to those of the In and Ga alloys, but in the case of Al and Ga the stability of AlN is much greater than that of GaN, which will tend to form AlN clusters. The growth of different alloy compositions, particularly the growth of higher In content GaInN, has stimulated some research on alternative precursors for nitrogen, as ammonia pyrolysis is not very efficient at temperatures below 800 °C. One of the favorite candidates is dimethylhydrazine, which will react readily with TMGa at temperatures down to 400 °C.

The reaction of the ammonia with the Group III alkyls to form adducts that can then polymerize is a problem associated with the high growth temperature. This requires special care over the introduction of the precursors, the control of gas flows and wall temperatures. A failure to adequately control these parasitic reactions will lead to poor growth efficiency, higher defect concentration in the GaN layer and poor dopant control. The dopants used for n-type and p-type GaN are Si from silane and Mg from dicyclopentadienylmagnesium. The n-type doping has proved to be fairly straightforward, but Mg doping results in the formation of Mg–H bonds that passivate the acceptor state. This problem was solved by annealing the epitaxial films after growth to remove the hydrogen. This is possible due to the thermal stability of GaN and the high mobility of hydrogen in the lattice. A further problem with p-type doping is that the Mg acceptor has an ionization energy of between 160 and 250 meV and only about 10% of the chemically introduced Mg is ionized at room temperature.

Despite the materials challenges of GaN and its alloys, **MOCVD** has enabled the production of a wide range of devices based on these alloys over the past decade, from high-power transistors to laser diodes. Both of these examples have required improvements in material quality and a reduction in the relatively high dislocation densities. In fact, the potential for nitrides is enormous as the quaternary GaInNAs can be tuned to around 1 eV with just 4% nitrogen and is a challenge to the use of InP-based materials for 1.3 μm telecommunications lasers.

14.2.6 II–VI MOCVD

The **MOCVD** of II–VI semiconductors is carried out at much lower temperatures than for their III–V counterparts and this has stimulated a wide range of research on alternative precursors, growth kinetics and energy-assisted growth techniques such as pho-

toassisted growth. The basic principles are the same as for III–V **MOCVD** and, in general, the same reaction chambers can be used but the lower growth temperatures have led to the development of new precursors, particularly for the Group VI elements. Hydrides are, in general, not used now but early work on ZnSe and ZnS used hydrogen selenide and hydrogen sulfide [14.50]. A strong prereaction occurred between the hydrides and dimethylzinc that could result in deposition at room temperature, but as with III–V **MOCVD**, prereactions can make it difficult to control the defect chemistry and the doping. These II–VI compounds and their alloys have been investigated as blue emitter materials with similar bandgaps to GaInN. Alternatively, ZnTe is a potential green emitter and the narrower bandgap tellurides are used for infrared detectors. In fact, the only commercial application of II–VI **MOCVD** has been for the fabrication of HgCdTe alloys for infrared detectors. However, the processes used are quite different to standard **MOCVD** and require different designs of reactor cells, as will be shown in the next section.

MOCVD of HgCdTe

HgCdTe is one of the few direct bandgap semiconductors suitable for infrared detection in the important 10 μm band. The alloy has only a 0.3% mismatch over the entire composition range and will cover the entire infrared spectrum from the near-infrared with CdTe to the far-infrared (HgTe is a semimetal so there is no lower limit to the band gap). The main difficulty with growing HgCdTe by **MOCVD** has been the very high equilibrium vapor pressure of Hg over the alloy even at relatively low temperatures. For example, **MBE** has to be carried out at temperatures below 200 °C. A further difficulty created by the instability of HgTe is that the tellurium-rich phase boundary, which represents the minimum Hg pressure required to achieve growth, has a high concentration of doubly ionized metal vacancies that make the material p-type. At typical **MOCVD** growth temperatures for HgCdTe, 350 to 400 °C, the equilibrium vapor pressure for Hg would have to be close to the saturated vapor pressure for liquid Hg in order to keep the metal vacancy concentration below the impurity background. This is clearly not realistic in **MOCVD** as the walls of the reaction chamber would have to be heated to the same temperature as the substrate to avoid mercury condensation, and this, in turn, would cause pyrolysis of the precursors before they arrived at the substrate. Fortunately, it is possible to grow HgCdTe film on the tellurium-rich phase boundary where the Hg

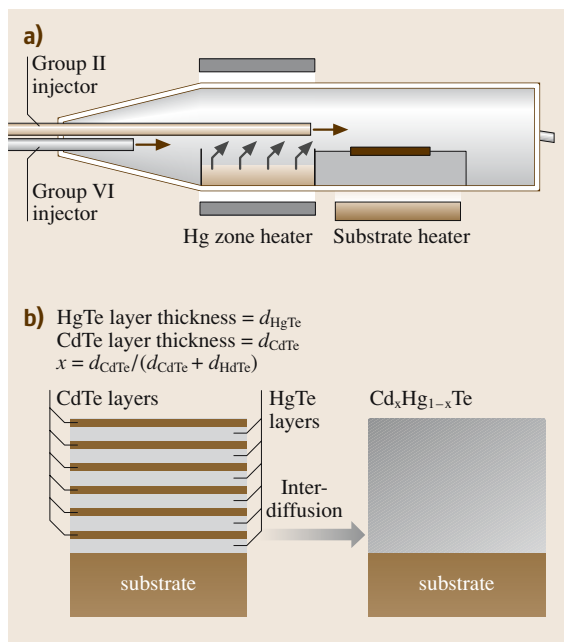


Fig. 14.16 (a) Schematic of MOCVD cell for HgCdTe growth, (b) schematic of IMP

source and reactor walls can be held at around 250 °C. At this temperature the pyrolysis of the Cd source, dimethylcadmium (DMCd), and the tellurium source, diisopropyltellurium (DIPTe), are sufficiently slow to minimize the premature reaction.

A schematic of an MOCVD reactor suitable for the growth of HgCdTe is shown in Fig. 14.16. The main features are the heating of the reactor wall and the internal source of liquid Hg. This is the only MOCVD process that uses an elemental source and is only possible because of the high vapor pressure of liquid Hg. In fact, a higher partial pressure can be achieved using the elemental source than an organometallic source, and this is the only metal where this is true. Another feature of this reactor cell is the injection of DMCd through the Hg source to avoid a radical exchange reaction between DMCd and Hg [14.51].

The alloy composition needs to be precisely controlled in order to control the detection wavelength of the infrared detector. For a 10 μm detector the proportion of Cd on the metal sub-lattice is 21% and it needs to be controlled to better than 0.5%. The reaction rates for CdTe and HgTe require different optimum flow rates, which makes simultaneous alloy control very difficult. This problem was overcome using the interdiffused multilayer process (IMP), where alternate layers

of HgTe and CdTe were grown and the flow rate optimized for each [14.52]. This process relied on very rapid Cd/Hg interdiffusion in the HgCdTe alloy, which enables complete homogenization of the alloy at the growth temperature. The composition is now simply controlled by the relative thicknesses of HgTe and CdTe, as shown schematically in Fig. 14.16b.

Sulfides and Selenides

Low-temperature growth of ZnSe (below 400 °C) has been achieved using ditertiarybutylselenide (DTBSe) with dimethylzinc (DMZn) or the adduct DMZn.TEN. The amine, triethylamine (TEN), was first introduced to suppress gas-phase reactions with the hydrides H₂Se and H₂S. However, this adduct has the additional advantage of reducing the saturated vapor pressure of DMZn and making it easier to manage. The significance of keeping the growth temperature low is to avoid deep-level native defects that act as trapping sites for donors and acceptors. The p-type doping of ZnSe with nitrogen was problematic due to hydrogen passivating the dopant. Incorporation of nitrogen from a variety of precursor sources such as amines and azides up to concentrations of 10¹⁸ cm⁻³ could be readily achieved but the active dopant concentration, in general, remained below 10¹⁵ cm⁻³. This was in contrast with the success of MBE growth of ZnSe that was doped from a nitrogen-plasma source. Fujita and Fujita [14.53] overcame this problem by ex situ annealing in a similar manner to GaN but the weaker lattice gives less scope for this treatment compared with GaN.

An alternative approach for low-temperature growth of ZnSe was to use short-wavelength light and UV wavelengths to illuminate the growing surface and promote surface photocatalytic reactions [14.54]. It was shown by Irvine et al. [14.55] that the reaction kinetics did not depend on the Se precursor, giving similar growth rates for DMSe, DESe and DIPSe, but depended on a hydrogen radical reaction that was initiated by a surface decomposition of the Group II precursor. Although this was effective for growing epitaxial films of ZnSe at temperatures well below 400 °C, it was clear that hydrogen incorporation was a natural consequence of the reaction mechanism.

MOCVD of Group II Oxides

A recent resurgence of interest in ZnO and related materials such as ZnMgO and ZnCdO has arisen because of the success of GaInN as a blue emitter and the potential for further developments with UV laser diodes and a solid state replacement for domestic lighting. GaInN white

light LEDs already exist but the efficiency of the phosphors would improve if they were excited with UV rather than blue photons. A further potential advantage of ZnO is that large ZnO single-crystal substrates can be grown by the hydrothermal method and would eventually avoid the defect problems associated with heteroepitaxy that have slowed progress with GaN.

All the early work on ZnO MOCVD used oxygen or water vapor as the oxygen source. These react strongly at room temperature with DMZn and DEZn. Although reasonably good quality ZnO films have been deposited with this approach, it is unlikely that it will lead to high-quality epitaxial growth or good doping control. Essentially, prereaction in all of the III–V and II–VI semiconductors has been a barrier to obtaining device-quality material. The favored alternative oxygen precursors are the alcohols: isopropanol and tertiarybutanol. For higher temperatures, N₂O is a suitable precursor. In general, for epitaxial growth on sapphire or ZnO substrates it is necessary to grow at temperatures above 600 °C, but for polycrystalline transparent conducting oxides (TCOs) these precursors can react at temperatures as low as 300 °C. It is possible to readily dope ZnO n-type using TMAI, but as with ZnSe it has been difficult to achieve p-type doping. Some encouraging results have been obtained using ammonia [14.56], but this work is still at an early stage of development and must be solved before electroluminescent devices can be made. This is proving

to be another class of materials where the versatility of MOCVD has a lot of potential for innovative solutions.

14.2.7 Conclusions

This section of the chapter has covered the basic principles of MOCVD and reviewed the range of III–V and II–VI semiconductors that can be grown in this way. This can be contrasted with LPE and MBE, where each method will have its own strengths and weaknesses for a particular material or application. The strength and the weakness of MOCVD is in its complexity. With the right precursors it is possible to deposit almost any inorganic material, but in many cases the reaction mechanisms are not well understood and the development is empirical, with the researcher spoiled by a very wide choice. This is not to deny the very considerable successes that have led to major industries in compound semiconductors that has been epitomized in the past 10 years by the production of GaN and the plethora of large LED displays that would not have been possible without MOCVD. Without the pioneering work of *Manasevit* and *Simpson*, who demonstrated the potential to grow so many of these materials in the early years, and the fortuitous ease with which GaAs/AlGaAs could be grown, we might not have tried so hard with the more difficult materials and hopefully we will see many more innovations in the future with MOCVD.

14.3 Molecular Beam Epitaxy (MBE)

14.3.1 Introduction and Background

MBE is conceptually a very simple route to epitaxial growth, in spite of the technology required, and it is this simplicity that makes MBE such a powerful technique. It can be thought of as a refined form of vacuum evaporation, in which neutral atomic and molecular beams from elemental effusion sources impinge with thermal velocities on a heated substrate under ultrahigh vacuum (UHV). Because there are no interactions within or between the beams, only the beam fluxes and the surface reactions influence growth, giving unparalleled control and reproducibility. Using MBE, complex structures can be grown atomic layer by atomic layer, with precise control over thickness, alloy composition and intentional impurity (doping) level. UHV confers two further advantages: cleanliness, because the partial pres-

sures of impurities are so low, and compatibility with in situ analytical techniques – essential to understanding the surface reaction kinetics. The basic elements of an MBE system are shown schematically in Fig. 14.17. A number of reviews [14.57–59] and books [14.60, 61] have discussed the physics, chemistry, technology and applications of MBE.

The technique that became known as MBE evolved from surface kinetic studies of the interaction of silane (SiH₄) beams with Si [14.62] and of Ga and As₂ beams with GaAs [14.63]. *Cho* and coworkers, who first used the term molecular beam epitaxy, demonstrated that MBE was a viable technique for the growth of III–V material for devices, leading the way for a worldwide expansion of effort.

Much early MBE equipment had a single vacuum chamber for loading, deposition and analysis, which

led to prolonged system pumpdown between growths. The technique required significant improvements in vacuum conditions before very high quality thin films were grown. *Cho* surrounded the effusion cells with a liquid nitrogen cryopanel to give thermal isolation and reported the MBE growth of thin films of n- and p-type GaAs for device purposes [14.64] and of GaAs/AlGaAs heterostructures [14.65]. The introduction of a substrate-exchange load lock [14.66] drastically reduced pumpdown times and reduced contamination of the deposition chamber. The installation of extensive internal liquid nitrogen-cooled cryopanel [14.67] substantially increased the pumping of oxygen containing species and permitted the growth of AlGaAs with superior quality. *Tsang* [14.68] demonstrated lasers with threshold current densities superior to those grown by LPE. Uniform growth over a 2"-diameter wafer was achieved by the introduction of a rotating substrate holder capable of 5 rpm [14.69]. Advanced forms of these features are now standard on commercial MBE systems, many of which feature a modular design. The technology of MBE is now mature, with increasing numbers of ever larger high-throughput, multiwafer production MBE machines in widespread use since the early 1990s.

MBE has been used to grow a wide range of materials, including semiconductors, superconductors, metals, oxides, nitrides and organic films. In almost all cases there is a drive to produce structures with ever smaller dimensions, whether for higher-performance devices, quantum confinement or, more recently, nanotechnology. This is longest established in III-V semiconductors, from the GaAs/AlGaAs superlattice [14.70] through quantum wells (QWs) and modulation doping to quantum wires and quantum dots [14.71]. Such low-dimensional structures form the basis of the QW lasers and p-HEMTs produced in huge volumes by MBE for optoelectronic and microwave applications. The combination of precise growth control and in situ analysis makes MBE the preeminent technology used to meet such demands.

14.3.2 Reaction Mechanisms

One of the first and most important of the in situ analytical techniques to be used in MBE is reflection high-energy electron diffraction (RHEED). From an early stage [14.72] it was used to determine the surface structure of the clean substrate and growing layer. This revealed that, in general, all surfaces are reconstructed (they have a lower symmetry than the bulk),

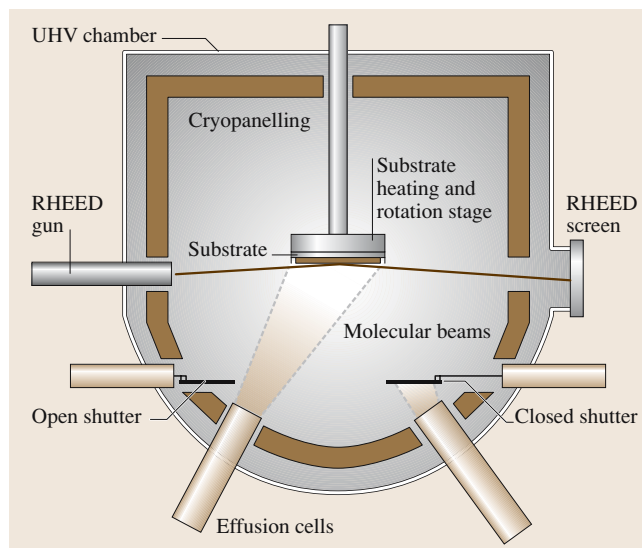


Fig. 14.17 Schematic diagram of an MBE growth chamber, showing the effusion cells and shutters, the substrate stage and the arrangement of the RHEED system

and *Cho* was the first to propose that the two-fold periodicity observed in the $[110]$ direction on the (001) surface was the result of dimerization of As atoms on the arsenic-terminated surface, which was confirmed many years later by scanning tunneling microscopy [14.73]. RHEED is a forward-scattering technique and therefore more compatible with the MBE arrangement of normally incident fluxes than the back-scattering geometry of LEED.

A further application of RHEED is the in situ measurement of growth rate. It was found [14.74,75] that the intensity of any diffraction feature oscillated with a period corresponding to the growth of a single monolayer (ML) – a layer of Ga + As – in the $[001]$ direction on a (001) substrate. These oscillations arise from surface morphological changes during two-dimensional (2-D) layer-by-layer growth, the Frank–van der Merwe mode, and a typical result for GaAs is shown in Fig. 14.18. The exact origin of the oscillations is still the subject of debate [14.59], but the technique was found to be applicable to many other material systems, including elemental semiconductors, metals, insulators, superconductors and even organic compounds.

Thus RHEED provided information on surface reconstruction and quantitative measurements of growth dynamics, which could be combined with those from other in situ techniques and related to theoretical treatments. In the early 1970s, the surface chemistry involved

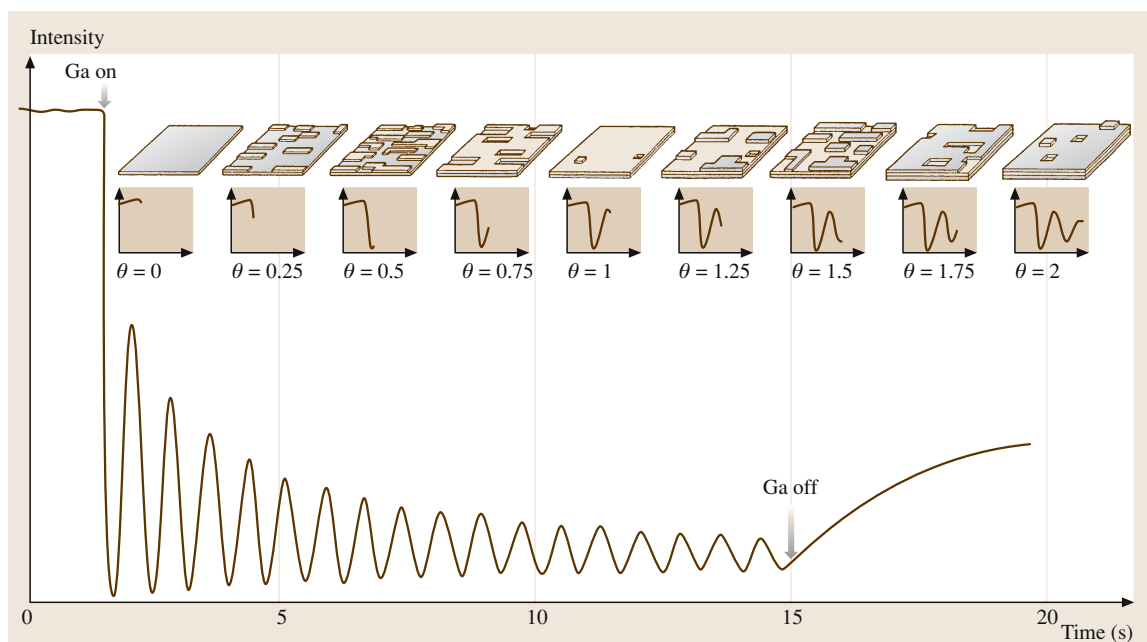


Fig. 14.18 Diagram illustrating the development of RHEED intensity oscillations during 2-D layer-by-layer growth, together with an actual experimental plot (courtesy of B. A. Joyce)

in the MBE growth of III-V compounds was studied extensively using a combination of RHEED and modulated molecular beam mass spectrometry [14.76] or temperature-programmed desorption [14.77].

Detailed information was obtained on surface reactions involving gallium and arsenic. It was found that

the dissociation of GaAs results in the desorption of As_2 , and not of As or As_4 [14.78], and that a significant amount of desorption takes place at MBE growth temperatures. The sublimation of elemental arsenic, as from an effusion cell, results in the formation of As_4 alone. It was also shown that growth from $\text{Ga} + \text{As}_2$ is a first-order reaction (Fig. 14.19), whereas growth from $\text{Ga} + \text{As}_4$ is second-order [14.79, 80]. These results suggest that in MBE growth Ga sticks to available As sites and chemisorption of As_2 occurs on available Ga atoms. Chemisorption of As_4 occurs with two As_4 molecules interacting on adjacent Ga atoms. The sticking coefficient of As_4 is observed to be less than or equal to 0.5, whereas the sticking coefficient of As_2 can be equal to one, in agreement with this model. This implies that maximum coverage will be less than 100%, since single Ga sites cannot be occupied, and that As vacancies will be introduced into material grown using As_4 . This was thought to be responsible for the higher deep-level concentrations observed, which were found to be reduced when As_2 was used or at higher growth temperatures. It is now known that above 580°C As_4 dissociates to As_2 on the surface [14.81].

Detailed models were established for the growth of GaAs [14.79, 80] and AlAs [14.82]. With minor modifications these models are also valid for ternary alloys

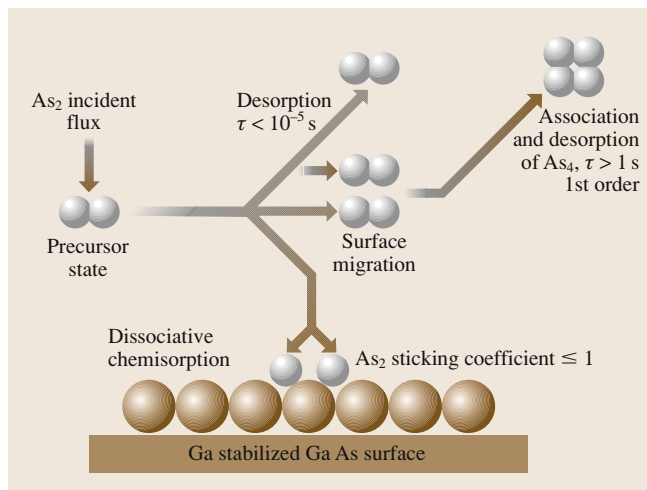


Fig. 14.19 Surface chemistry of the growth of GaAs from As_2 and Ga (courtesy of B. A. Joyce)

such as AlGaAs [14.83]. It was established that growth rate depends solely on the net Group III flux (incident flux minus desorbing flux), and that ternary alloy composition can be controlled by adjusting the ratio of the Group III fluxes provided the thermal stabilities of both of the binary compounds that make up the alloy are considered. The Group V element need only be supplied in excess. The situation is more complicated for alloys containing both arsenic and phosphorus, since the presence of one Group V element influences the sticking coefficient of the other. *Foxon et al.* [14.84] found that phosphorus has a much lower incorporation probability than arsenic.

More recent studies have made use of in situ STM and more powerful theoretical treatments to consider nucleation and growth at the atomistic level, but the basic models are still sound.

14.3.3 MBE Growth Systems

The UHV system required for MBE is of conventional stainless steel construction, with an ultimate or background vacuum of less than 5×10^{-11} torr achievable with a clean system after baking, and with the liquid nitrogen cryopanels filled. At such a pressure the molecules have a typical mean free path of 10^6 m and so only suffer collisions with the internal surfaces of the system. It would take several hours to build up a monolayer of impurity on the wafer surface.

Oil-free pumping is used to eliminate the possibility of contamination by hydrocarbon backstreaming; typically rough pumping is with sorption pumps and UHV pumping is with ion pumps and titanium sublimation pumps. Diffusion pumps or turbomolecular pumps can be used, however, provided suitable cold traps are fitted, and such pumps are required for the higher gas loads involved in GSMBE and CBE. A two- or three-stage substrate entry load-lock and preparation chamber isolated by gate valves is used to minimize the exposure of the system to air. In modular systems, further deposition and analysis chambers may be added to the system and samples are transferred via the preparation chamber under UHV conditions.

The growth chamber substrate stage is surrounded by a large liquid nitrogen-cooled cryopanel, which has a high pumping speed for H_2O , CO , O_2 and other condensable species. This is arranged so that the heated (500 – 700°C) substrate is not directly exposed to thermal sources other than the molecular beams themselves, and impurities emanating from any other source can only reach the substrate after suffering at least one colli-

sion with, and probable adsorption by, a surface at liquid nitrogen temperatures. The cryopanel also reduces contamination arising from outgassing from the walls of the chamber that are exposed to radiation from the effusion cells.

Control over the composition and doping levels of the epitaxial layers is achieved by precise temperature control of the effusion cells and the use of fast-acting shutters in front of these cells. In most systems, the growth rate is about 1 monolayer per second, and the shutter operation time of 0.1 s thus corresponds to less than a monolayer of growth. A growth rate of one monolayer per second closely approximates one micrometer per hour for GaAs/AlGaAs.

Although often referred to as K-cells, the solid-source effusion cells used in MBE growth have a large orifice so as to obtain a high flux at reasonable temperatures. A true Knudsen cell has a very small outlet orifice compared to the evaporating surface, so that an equilibrium vapor pressure, typically 10^{-3} torr, is maintained within the cell. The diameter of this orifice is less than one tenth of the molecular mean free path, which is typically several centimeters. Under these conditions, a near approximation to ideal Knudsen effusion is obtained from the cell, giving molecular flow with an approximately cosine distribution. The flux from such a cell can be calculated quite accurately, but a high temperature is required to produce a reasonable growth rate; for example a gallium Knudsen cell would need to be at 1500°C to produce the same flux as an open-ended effusion cell at 1000°C . The lower operating temperature helps to reduce impurities in the flux and puts a lower thermal load on the system. The beam from an open-ended cell may not be calculable with any degree of accuracy but it is highly reproducible. Once calibrated via growth rate, normally by in situ measurement, the flux can be monitored using an ion gauge located on the substrate stage.

A number of effusion cells can be fitted to the growth chamber, generally in a ring facing towards the substrate with the axis of each cell at an angle of 20 – 25° to the substrate normal. Simple geometrical considerations therefore dictate the best possible uniformity that can be achieved with a stationary substrate [14.85]. Associated with each cell is a fast-action refractory metal shutter with either pneumatic or solenoid operation.

Uniformity of growth rate for a binary compound can be achieved by rotating the substrate at speeds as low as a few rpm, but compositional uniformity of ternary or quaternary alloys requires rotation of the substrate at speeds of up to 120 rpm (normally rotation is timed

so that one rotation corresponds to the growth of one monolayer). The mechanical requirements for a rotating substrate stage in a UHV system are quite demanding, as no conventional lubricants can be used on the bearings or feedthroughs and yet lifetime must exceed several million rotations. Magnetic rotary feedthroughs have largely replaced the earlier bellows type.

The need to rotate the substrate to give uniformity also leads to complications in substrate temperature measurement. The substrate is heated by radiation from a set of resistively heated tantalum foils behind the substrate holder, and both the heater and the thermocouple are stationary. Without direct contact between the thermocouple and the wafer the “indicated” thermocouple temperature will be very different from the “actual” substrate temperature. Some form of calibration can be obtained by using a pyrometer, although problems with window coating, emissivity changes and substrate transparency below the bandgap impose limits on the accuracy of such measurements. Alternatively, a number of “absolute” temperature measurements can be obtained by observing transitions in the **RHEED** pattern, which occur at reasonably fixed temperatures. However, such transitions occur in the lower temperature range and extrapolation to higher growth temperatures is not completely reliable. If the substrate is indium bonded to a molybdenum block (using the surface tension of the indium to hold the substrate), then inconsistencies in wetting can lead to variations in temperature across the substrate. Most modern systems and all production machines use “indium-free” mounting, which avoids these problems. However, the substrate is transparent to much of the **IR** radiation from the heater, putting a higher thermal load on the system.

The substrate preparation techniques used prior to **MBE** growth are very important, as impurities on the surface provide nucleation sites for defects. Historically, various chemical clean and etch processes were used, but wafers are now usually supplied “epi-ready”, with a volatile oxide film on the surface that protects the surface from contamination and can be thermally removed within the UHV chamber. **RHEED** is used to confirm the cleanliness of the surface prior to growth.

Historically, one of the major problems in **MBE** was the presence of macroscopic defects, with a typical density of 10^3 – 10^5 cm⁻², although densities below 300 cm⁻² were reported for ultraclean systems [14.86]. Defects are generally divided into two types; small hillocks or pits and oval defects. Such defects are a serious obstacle to the growth of material for integrated circuits, and considerable effort was devoted to the

problem. Oval defects are microtwin defects originating at a local imperfection, oriented in the (110) direction and typically 1 to 10 μm in length [14.87]. There are several possible sources of these defects, including foreign impurities on the substrate surface due to inadequate substrate preparation or to oxides from within the system, and possibly from the arsenic charge or the condensate on the cryopanel. The fact that oval defects were not seen when graphite crucibles were used but were common with PBN crucibles suggests that gallium oxide from the gallium melt is a major source of such defects since oxides would be reduced by the graphite. *Chai and Chow* [14.88] demonstrated a significant reduction in defects by careful charging of the gallium source and prolonged baking of the system. The irregular hillocks and pits seen in **MBE**-grown material were probably produced by microdroplets of gallium spitting from the effusion cells on to the substrate surface [14.89]. Gallium spitting can be caused by droplets of Ga that condense at the mouth of the effusion cell, fall back into the melt and explode, ejecting droplets of liquid Ga, or by turbulence in the Ga melt due to uneven heating that causes a sudden release of vapor and droplets.

Continuous developments in the design of Group III effusion cells for solid-source **MBE** have largely eliminated the problem of macroscopic defects. Large-area Ta foil K-cell heaters have reduced the uneven heating of the PBN crucible; the use of a “hot-lipped” or two-temperature Group III cell, designed with a high-temperature front end to eliminate the condensation of gallium metal at the mouth of the cell, significantly reduced the spitting of microdroplets. Combined with careful procedures and the use of an arsenic cracker cell, defect densities as low as 10 cm⁻² have been reported [14.90].

There have also been developments in the design of Group V cells. A conventional arsenic effusion cell produces a flux of As₄, but the use of a thermal cracker to produce an As₂ flux resulted in the growth of GaAs with better optical properties and lower deep-level concentrations [14.91]. In the case of phosphorus, growth from P₂ was strongly preferred to that from P₄ for several reasons [14.92], and this was one reason behind the development of **GSMBE** described below. However, the use of phosphine requires suitable pumping and safety systems. The Group V cracker cell has two distinct zones. The first comprises the As or P reservoir and produces a controlled flux of the tetramer; this passes through the second – high-temperature – zone, where dissociation to the dimer occurs. Commercial

high-capacity cracker cells, some including a valve between the two zones to allow fast switching of Group V flux, have been developed for arsenic and phosphorus.

Commercial MBE systems have increased throughput with multiwafer substrate holders, cassette loading and UHV storage and preparation chambers linked to the growth chamber with automated transfer, while increased capacity effusion cells have reduced the downtime required for charging. In some cases additional analytical and processing chambers have been added to permit all-UHV processing of the device structure.

14.3.4 Gas Sources in MBE

A number of MBE hybrids were developed that combined the advantages of UHV deposition and external gas or metalorganic sources to produce a versatile technique that has some advantages over MBE and MOVPE. These techniques utilize the growth chambers developed for MBE and pumping systems with a high continuous throughput, typically liquid nitrogen-trapped diffusion pumps or turbomolecular pumps. Layers are deposited from molecular beams of the precursor materials introduced via gas source cells that are essentially very fine leak controllers. As in MBE, there are no interactions within or between beams and the precise control of beams using fast-acting gas-line valves is therefore translated into precise control of the species arriving at the substrate. Shutters are not generally required; atomically sharp interfaces and monolayer structures can be defined as a consequence of submonolayer valve switching times.

Panish [14.93] investigated the use of cracked arsine and phosphine in the epitaxial growth of GaAs and InP, later extending this work to the growth of GaInAsP, and suggested the name gas source MBE (GSMBE). The major advantage of gaseous Group V sources was that the cracker cells produced controllable fluxes of the dimers As₂ and P₂, giving improved control of the As:P ratio. The MBE growth of high-quality GaAs from cracked arsine and elemental gallium was demonstrated at the same time by Calawa [14.94].

The extension to gaseous Group III sources was made by Veuhoff et al. [14.95], who investigated the MOCVD of GaAs in a simple MBE system using trimethylgallium (TMGa) and uncracked arsine. Further study showed that cracking of arsine at the substrate surface was negligible, leading to the conclusion that unintentional cracking of the arsine had taken place in the inlet system. The acronym metalorganic MBE (MOMBE) was used to describe GaAs growth from

TMGa and cracked arsine in a modified commercial MBE system [14.96].

The growth of device-quality GaAs, InGaAs and InP from alkyl sources of both Group III and Group V elements was demonstrated by Tsang [14.97], who used the alternative acronym chemical beam epitaxy (CBE). The use of Group V alkyls, which had much poorer purity than the hydrides, was undertaken for safety reasons [14.98]. Material quality was improved when cracked arsine and phosphine were used [14.99]. RHEED observations indicated that reconstructed semiconductor surfaces could be produced prior to growth, as for MBE [14.100], and GaAs/AlGaAs quantum well structures were demonstrated that were comparable with those grown by MBE or MOVPE [14.101].

Almost all of this work was with III–Vs, where MOMBE/CBE was seen to have several significant advantages over MBE while retaining many of its strengths, including in situ diagnostics. The use of vapor Group III sources would avoid the morphological defects associated with effusion cells, and higher growth rate and greater throughput could be achieved. Both Group III and Group V sources were external, allowing for easy replacement without the need to break vacuum. Flux control with mass flow controllers (MFCs) and valves would improve control over changes in composition or doping level, since flow could be changed faster than effusion cell temperature. Abrupt changes could thus be achieved that would require switching between two preset effusion cells in MBE (a problem when the number of cells was limited by geometry). It also offered improved long-term flux stability and greater precursor flexibility. As this was still a molecular beam technique, precise control over layer growth and abrupt interfaces would be retained, without any of the gas phase reactions, boundary layer problems or depletion of reagents associated with MOVPE. Other advantages included improved InP quality using a P₂ flux, lower growth temperatures and selective-area epitaxy.

There was, however, a price to pay in system complexity, with the need for gas handling and high-volume pumping arrangements added to the expensive UHV growth chamber. These would have been acceptable if CBE had demonstrated clear advantages, but there were a number of other issues. The standard Al and Ga precursors used in MOVPE (trimethylaluminum and trimethylgallium) produced strongly p-type material when used in CBE, due to the incorporation of C as an acceptor. Triethylgallium proved to be a viable Ga source, but alloy growth was more complicated; no universally acceptable Al source was found, while InGaAs

growth was found to be strongly temperature-dependent. The surface chemistry associated with metalorganic sources proved complex and the temperature dependence of surface reactions not only restricted growth conditions, but also had a serious impact on uniformity and reproducibility, particularly for quaternary alloys such as GaInAsP [14.102]. The lack of suitable gaseous dopant sources, particularly for Si, was a further handicap [14.103], but the deliberate use of C for p-type doping proved a success and this was transferred to MBE and MOVPE. Carbon diffuses significantly less than the 'standard' MBE and MOVPE dopants, Be and Zn, respectively [14.104], and proved an ideal dopant for thin highly doped layers such as the base region in heterojunction bipolar transistors (HBTs) and for p-type Bragg reflector stacks in vertical cavity surface-emitting laser structures (VCSELs).

GSMBE remains important, not in the III–V field where, with some exceptions [14.105], the development of high-capacity Group V cracker cells provided an easier route to an As₂ and P₂ flux, but in the III–nitride field. There are two major routes to nitride MBE: active nitrogen can be supplied by cracking N₂ in an RF or ECR plasma cell, or ammonia can be injected and allowed to dissociate on the substrate surface. In contrast, CBE has not demonstrated a sufficient advantage over its parent technologies to be commercially successful, particularly as both MOVPE and MBE have continued to develop as production techniques.

14.3.5 Growth of III–V Materials by MBE

Although they were amongst the earliest materials to be grown by MBE, GaAs-based alloys retain great importance, with MBE supplying materials for the mass production of optoelectronic and microwave devices and leading research into new structures and devices.

GaAs/AlGaAs

AlGaAs is an ideal material for heterostructures, since AlAs has a greater bandgap than GaAs and the two have negligible mismatch ($\approx 0.001\%$). The growth rate in MBE depends on the net Group III flux, with one micrometer per hour corresponding to a flux of 6.25×10^{14} Ga (or Al) atoms $\text{cm}^{-2}\text{s}^{-1}$. At low growth temperatures, all incident Group III atoms are incorporated into the growing film, together with sufficient arsenic atoms to maintain stoichiometry, and excess arsenic atoms are desorbed. However, III–V compounds are thermally unstable at high temperatures. Above ≈ 600 K [14.79] arsenic is preferentially desorbed, so an excess arsenic

flux is required to maintain stoichiometry. At higher temperatures, loss of the Group III element becomes significant, so that the growth rate is less than would be expected for the incident flux. This is particularly important for the growth of AlGaAs, where growth temperatures above 650°C are generally used to give the best optical properties. The Ga flux must be significantly increased above that used at lower temperatures in order to maintain the required composition of the alloy.

Typical growth temperatures for MBE of GaAs are in the range $580\text{--}650^\circ\text{C}$ and material with high purity and low deep-level concentrations has been obtained in this temperature range [14.106, 107]. The commonly used dopants, Be (p-type) and Si (n-type), show excellent incorporation behavior and electrical activity at these temperatures and at moderate doping levels. As was noted above, for highly doped layers Be has largely been replaced by C, which diffuses somewhat more slowly. At doping levels above $\approx 5 \times 10^{17}$, Si occupies both Ga (donor) and As (acceptor) sites, producing electrically compensated material with a consequent reduction in mobility. It is still predominantly a donor, however, and is the best available n-type dopant. The electrical properties of GaAs also depend on the As/Ga flux ratio, since this influences the site occupancy of dopants. The optimum As/Ga ratio is that which just maintains As-stabilized growth conditions, which can be determined using RHEED observations of surface reconstruction.

MBE-grown GaAs is normally p-type, the dominant impurity being carbon [14.108]. The carbon concentration was found to correlate with CO partial pressure during growth [14.109] – CO is a common background species in UHV, being synthesized at hot filaments. The lowest acceptor levels commonly achieved are of the order of $5 \times 10^{13} \text{ cm}^{-3}$, and such layers can be lightly doped to give n-type material with high mobilities. However, very high purity GaAs has been produced by adjusting the operating conditions for an arsenic cracker cell [14.110], which suggests that carbon contamination originates from hydrocarbons in the As charge. Unintentionally doped GaAs was n-type with a total impurity concentration of $< 5 \times 10^{13} \text{ cm}^{-3}$ and a peak mobility of $4 \times 10^5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at 40 K, the highest reported for n-type GaAs. The 77 K mobility of $> 200\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ is comparable with that for the highest purity GaAs grown by LPE [14.111].

MBE is capable of the growth of very high-quality material for structures whose physical dimensions are comparable to the wavelength of an electron (or hole) so that quantum size effects are important. Such structures have typical layer thicknesses from 100 \AA down to

2.8 Å, the thickness of a monolayer of GaAs, and have doping and composition profiles defined on an atomic scale. These low-dimensional structures have become very important in III–V device technology, one example being the modulation-doped heterojunction.

The precise control of growth that is possible is demonstrated by the very high mobilities obtained for modulation-doped structures grown by MBE. The modulation-doped GaAs/AlGaAs heterojunction has a band structure that causes carriers from the highly doped AlGaAs to be injected into a thin undoped region in the GaAs (Fig. 14.20a). Since the carriers are then separated from the donor atoms that normally scatter them, and confined in a quantum well as a two-dimensional electron gas (2DEG), very high mobilities can be achieved. Through suitable design of the structure and the use of lightly doped AlGaAs, a GaAs/AlGaAs 2DEG structure was produced with a peak mobility of $1.1 \times 10^7 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at 1.3 K and a 4 K mobility of $4.5 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and with a sheet carrier concentration of 10^{12} cm^{-2} [14.112]. GaAs layers grown in the same machine typically exhibited 77 K mobilities of the order of $100\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Similar results can be obtained by delta doping the AlGaAs, that is by confining the Si donor atoms to a single monolayer in the AlGaAs separated by a few nm from the well, a technique with a wide range of applications [14.113].

The major application of this structure for device purposes is the high electron mobility transistor (HEMT), also known as the modulation-doped FET (MODFET), an FET in which the carriers are confined to the two-dimensional layer (Fig. 14.20b). The main advantage of the HEMT is not the increase in mobility, which is modest at room temperature for a practical device, but the very low noise when it is operated as a microwave amplifier, due to the reduction in impurity scattering. The HEMT is an essential component of many microwave systems, including mobile phones and satellite TV receivers.

MBE became used as a production technology for GaAs-based devices because of the excellent uniformity and reproducibility possible, and because the extensive UHV load-lock system of a production machine permits a large number of runs to be undertaken on a continuous basis. This was demonstrated as early as 1991 by a number of manufacturers producing both HEMTs and GaAs/AlGaAs lasers [14.114]. By 1994 some 5 million MQW lasers for compact disc applications were produced per month by MBE. Other devices included HEMTs and InAs Hall sensors. A comparison of production costs made at this time [14.115] showed that

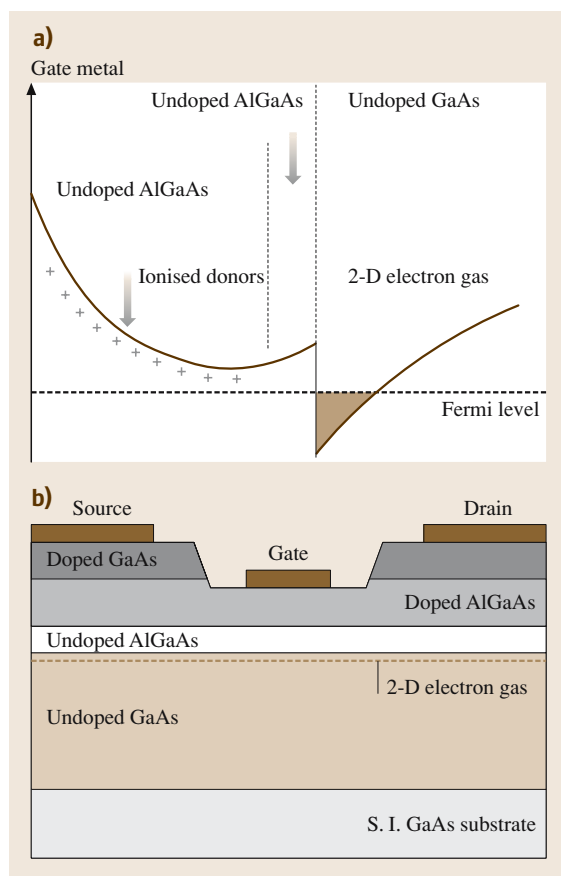


Fig. 14.20a,b Schematic diagrams of the band structure (a) and physical arrangement (b) of the high electron mobility transistor (HEMT)/modulation-doped FET

capital costs for MBE were very similar to those for MOVPE, which required costly safety systems for the hydride gases, while MBE costs per wafer were somewhat less than for MOVPE. The growth in demand for HEMTs for mobile phones drove a further expansion in production MBE, although MOVPE remained dominant in InP-based optoelectronic devices.

InGaAs

InGaAs is generally grown at lower temperatures than GaAs/AlGaAs because of the higher rate of In desorption. Both InGaAs and the quaternary InGaAsP can be lattice-matched to InP substrates; heterostructures in these materials form the basis of the 1.3 μm and 1.55 μm optoelectronic devices used in telecommunications. InAs has a lattice mismatch of $\approx 7\%$ with GaAs; thin films of InGaAs can be deposited pseudo-

morphically but as the In content is increased the critical thickness falls, so that little more than a **ML** of InAs can be grown in this way. Thicker films will be relaxed by the formation of dislocations, limiting their usefulness for devices. The pseudomorphic **HEMT** (or **pHEMT**) replaces the GaAs channel of the conventional **HEMT** with a strained InGaAs channel in order to take advantage of the greater carrier confinement and superior electron transport properties. The maximum In content of the channel is limited by the need to prevent relaxation.

Under certain growth conditions InAs or InGaAs islands are formed spontaneously. It is generally accepted that the growth of InAs on GaAs (001) follows a version of the Stranski–Krastanov mode, which implies that following the deposition of ≈ 1.7 **ML** of InAs in a 2-D pseudomorphic form (sometimes referred to as the wetting layer), coherent 3-D growth is initiated by a very small increment (≤ 0.1 **ML**) of deposited material to relax the elastic strain introduced by the lattice mismatch. The **QDs** rapidly reach a saturation number density, which is both temperature- and In flux-dependent, with a comparatively narrow size (volume) distribution. The actual process is rather more complicated, not least as a result of alloying with the GaAs substrate, and is the subject of much current research [14.59].

These islands can be embedded in a layer of GaAs to form self-assembled quantum dots (**QDs**), which have become a topic of immense interest due to the potential application of **QDs** in a wide range of devices, especially lasers. The volume fraction of **QDs** in an active layer can be increased by building up a 3-D array; the strain field induced around each dot influences not only the inter-dot spacing but also the capping layer growth, so that subsequent layers of dots are aligned ([14.71] and references therein). **QD** lasers offer a route to long-wavelength emission from GaAs-based devices [14.116].

Group III Nitrides

The growth of Group III nitrides has been dominated by **MOVPE** since the demonstration of a high-brightness blue-emitting InGa_N-AlGa_N double heterostructure **LED** by *Nichia* [14.117] and the subsequent development of other optoelectronic devices, including laser diodes also emitting in the blue [14.118]. **MBE** has made a significant contribution to more fundamental studies and to the growth of nitrides for high-power and microwave devices. Once again the wide range of in situ diagnostic techniques available has been important.

For heteroepitaxial growth on the most commonly used substrates, sapphire and silicon carbide, several parameters strongly influence the quality of material

produced. These include substrate cleaning, initial nitridation, the nucleation and coalescence of islands involved in the low-temperature growth of a buffer layer, and subsequent annealing at a higher temperature. The polarity (nature of the outermost layer of atoms) of {0001}-oriented hexagonal structure films also has a crucial influence on material quality, but both N- or Ga-polarity can occur with **MBE** growth on sapphire substrates [14.119]. Under typical growth conditions with **MOVPE**, however, Ga-polarity material is exclusively produced.

Several of these problems can be resolved using GaN templates obtained by growing thick layers onto suitable substrates using **MOVPE** and then exploiting the advantages of **MBE** to produce the functional layer on the GaN template. These advantages include well-controlled layer-by-layer growth and lower growth temperatures than those used for **MOVPE**, so that InGa_N phase separation and In desorption are less problematic and precise quantum wells can be grown. No post-growth thermal annealing is required to activate the p-type dopant. In this way films have been produced with smooth surface morphology and high performance, although **MBE** was still limited to low-power **LEDs** until the recent demonstration of laser diodes [14.120].

In the use of active nitrogen from plasma sources, the III/V flux ratio at the substrate during growth is also a critical parameter. GaN layers grown with a low III/V flux ratio (N-stable growth) display a faceted surface morphology and a tilted columnar structure with a high density of stacking faults. Smooth surfaces are only obtained under Ga-rich conditions, where not only is there a dramatic reduction in surface roughness, but significant improvements in structural and electrical properties are also observed. This is, of course, the exact opposite of the growth of most III–V compounds, such as GaAs. In the case of nitrides, it is thought that Ga-rich conditions (close to the point where Ga droplets are formed) promote step flow growth, whereas N-stable growth promotes the nucleation of new islands. In contrast, growth from NH₃ is smoother under N-rich conditions [14.121].

Group III–V Nitrides

The “dilute nitrides” are III–V–N materials such as GaAs_N and GaInNAs, where the N concentration is $\leq 2\%$. Replacing a small fraction of As atoms with smaller N atoms reduces both the lattice constant and the bandgap. Adjusting the composition of GaInNAs allows the bandgap, band alignment, lattice constant and strain to be tailored in a material that can be lattice-

matched with GaAs [14.122]. This offers strong carrier confinement and thermal stability compared to InP-based devices operating at 1.3 μm and 1.55 μm , and allows GaAs VCSEL technology to be exploited at these wavelengths [14.123].

These materials are grown in a metastable regime at a low growth temperature because of the miscibility gap in the alloys, so a less stable precursor than NH_3 is needed. A nitrogen plasma source provides active N without the incorporation of hydrogen during growth associated with hydride sources, thus avoiding the deleterious formation of N–H bonds. Material with excellent crystallinity and strong PL at 1.3 μm can be obtained by optimizing growth conditions and using post-growth rapid thermal annealing [14.124]. A number of challenges remain, particularly in the higher N material required for longer wavelengths, including the limited solubility of N in GaAs and nonradiative defects caused by ion damage from the N plasma source. These challenges may be met by using GaInNAsSb; the addition of Sb significantly improves the epitaxial growth

and the material properties, and enhanced luminescence is obtained at wavelengths longer than 1.3 μm [14.125].

14.3.6 Conclusions

MBE, historically seen as centered on GaAs-based electronic devices, has broadened its scope dramatically in both materials and devices. In addition to the materials described above, MBE has been used to grow epitaxial films of a wide range of semiconductors, including other III–V materials such as InGaAsP/InP and GaAsSb/InAsSb; silicon and silicon/germanium; II–VI materials such as ZnSe; dilute magnetic semiconductors such as GaAs:Mn [14.126] and other magnetic materials. It has also been used for the growth of metals, including epitaxial contacts for devices, oxides [14.127] and organic films [14.128]. Two clear advantages possessed by MBE are the wide range of analytical techniques compatible with a UHV system and the precise control of growth to less than a monolayer, which give it unrivaled ability to grow quantum dots and other nanostructures.

References

- 14.1 D. Elwell, H.J. Scheel: *Crystal Growth from High-Temperature Solutions* (Academic, New York 1975)
- 14.2 H. Nelson: RCA Rev. **24**, 603 (1963)
- 14.3 R. L. Moon: J. Cryst. Growth **170**, 1 (1997)
- 14.4 H. J. Scheel: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 14.5 P. Capper, T. Tung, L. Colombo: *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman & Hall, London 1997)
- 14.6 M. B. Panish, I. Hayashi, S. Sumski: Appl. Phys. Lett. **16**, 326 (1970)
- 14.7 M. G. Astles: *Liquid Phase Epitaxial Growth of III–V Compound Semiconductor Materials and their Device Applications* (IOP, Bristol 1990)
- 14.8 V. A. Dmitriev: Physica B **185**, 440 (1993)
- 14.9 T. Cizzek: *The Technology of Crystal Growth and Epitaxy*, ed. by H. J. Scheel, T. Fukuda (Wiley, Chichester 2003)
- 14.10 M. I. Alonso, K. Winer: Phys. Rev. B **39**, 10056 (1989)
- 14.11 V. A. Dmitriev: *Properties of Silicon Carbide*, EMIS Datareview Series, ed. by G. L. Harris (IEE, London 1995) p. 214
- 14.12 Y. Mao, A. Krier: Mater. Res. Soc. Symp. Proc **450**, 49 (1997)
- 14.13 A. Krier, Z. Labadi, A. Manniche: J. Phys. D: Appl. Phys. **32**, 2587 (1999)
- 14.14 M. Mauk: private communication (2004)
- 14.15 H. Yamane, M. Shimada, T. Sekiguchi, F. J. DiSalvo: J. Cryst. Growth **186**, 8 (1998)
- 14.16 C. Klemenz, H. J. Scheel: J. Cryst. Growth **211**, 62 (2000)
- 14.17 A. Krier, H. H. Gao, V. V. Sherstinov: IEE Proc. Optoelectron **147**, 217 (2000)
- 14.18 E. R. Rubstov, V. V. Kuznetsov, O. A. Lebedev: Inorg. Mater. **34**, 422 (1998)
- 14.19 M. G. Mauk, Z. A. Shellenbarger, P. E. Sims, W. Bloothoofd, J. B. McNeely, S. R. Collins, P. I. Rabinowitz, R. B. Hall, L. C. DiNetta, A. M. Barnett: J. Cryst. Growth **211**, 411 (2000)
- 14.20 J.-i. Nishizawa, K. Suto: *Widegap II–VI Compounds for Optoelectronic Applications*, ed. by H. E. Ruda (Chapman & Hall, London 1992)
- 14.21 F. Sakurai, M. Motozawa, K. Suto, J.-i. Nishizawa: J. Cryst. Growth **172**, 75 (1997)
- 14.22 M. G. Astles: *Properties of Narrow Gap Cadmium-Based Compounds*, EMIS Datareview series, ed. by P. Capper (IEE, London 1994) pp. 13, 380
- 14.23 B. Pellicciari, J. P. Chamonal, G. L. Destefanis, L. D. Cioccio: Proc. SPIE **865**, 22 (1987)
- 14.24 P. Belca, P. A. Wolff, R. L. Aggarwal, S. Y. Yuen: J. Vac. Sci. Technol. A **3**, 116 (1985)
- 14.25 S. H. Shin, J. Pasko, D. Lo: Mater. Res. Soc. Symp. Proc. **89**, 267 (1987)
- 14.26 A. Wasenczuk, A. F. M. Willoughby, P. Mackett, E. S. O'Keefe, P. Capper, C. D. Maxey: J. Cryst. Growth **159**, 1090 (1996)

- 14.27 T. Tung, L.V. DeArmond, R.F. Herald: Proc. SPIE **1735**, 109–134 (1992)
- 14.28 P.W. Norton, P. LoVecchio, G.N. Pultz: Proc. SPIE **2228**, 73 (1994)
- 14.29 P. Capper, J. Gower, C. Maxey, E. O'Keefe, J. Harris, L. Bartlett, S. Dean: *Growth and Processing of Electronic Materials*, ed. by N. McN. Alford (IOM Communications, London 1998)
- 14.30 C.C. Wang: J. Vac. Sci. Technol. B **9**, 1740 (1991)
- 14.31 T.N. Casselman, G.R. Chapman, K. Kosai, et al.: *U.S. Workshop on Physics and Chemistry of MCT and other II–VI compounds*, Dallas, TX (Oct. 1991)
- 14.32 R.S. Patrick, A.–B. Chen, A. Sher, M.A. Berding: J. Vac. Sci. Technol. A **6**, 2643 (1988)
- 14.33 A. Rogalski: *New Ternary Alloy Systems for Infrared Detectors* (SPIE, Bellingham 1994)
- 14.34 A.A. Chernov, H.J. Scheel: J. Cryst. Growth **149**, 187 (1996)
- 14.35 H.M. Manasevit, W.I. Simpson: J. Electrochem. Soc. **116**, 1725 (1969)
- 14.36 A.A. Chernov: Kinetic processes in vapor phase growth. In: *Handbook of Crystal Growth*, ed. by D.T.J. Hurle (Elsevier, Amsterdam 1994)
- 14.37 G.B. Stringfellow: J. Cryst. Growth **115**, 1 (1991)
- 14.38 D.M. Frigo, W.W. van Berkel, W.A.H. Maassen, G.P.M. van Mier, J.H. Wilkie, A.W. Gal: J. Cryst. Growth **124**, 99 (1992)
- 14.39 S. Tompa, M.A. McKee, C. Beckham, P.A. Zwadzki, J.M. Colabella, P.D. Reinert, K. Capuder, R.A. Stall, J.E. Norris: J. Cryst. Growth **93**, 220 (1988)
- 14.40 X. Zhang, I. Moerman, C. Sys, P. Demeester, J.A. Crawley, E.J. Thrush: J. Cryst. Growth **170**, 83 (1997)
- 14.41 P.M. Frijlink, J.L. Nicolas, P. Suchet: J. Cryst. Growth **107**, 166 (1991)
- 14.42 D.W. Kisker, J.N. Miller, G.B. Stringfellow: Appl. Phys. Lett. **40**, 614 (1982)
- 14.43 C.A. Larson, N.I. Buchan, S.H. Li, G.B. Stringfellow: J. Cryst. Growth **93**, 15 (1988)
- 14.44 M.W. Raynor, V.H. Houlding, H.H. Funke, R. Frye, J.A. Dietz: J. Cryst. Growth **248**, 77–81 (2003)
- 14.45 R.M. Biefeld, R.W. Gedgridge Jr.: J. Cryst. Growth **124**, 150 (1992)
- 14.46 C.A. Wang, S. Salim, K.F. Jensen, A.C. Jones: J. Cryst. Growth **170**, 55 (1997)
- 14.47 S. Nakamura: Jpn. J. Appl. Phys. **30**, 1620 (1991)
- 14.48 A. Stafford, S.J.C. Irvine, K. Jacobs. Bougrioua, I. Moerman, E.J. Thrush, L. Considine: J. Cryst. Growth **221**, 142 (2000)
- 14.49 S. Keller, S.P. DenBaars: J. Cryst. Growth **248**, 479 (2003)
- 14.50 B. Cockayne, P.J. Wright: J. Cryst. Growth **68**, 223 (1984)
- 14.51 W. Bell, J. Stevenson, D.J. Cole–Hamilton, J.E. Hails: Polyhedron **13**, 1253 (1994)
- 14.52 J. Tunnicliffe, S.J.C. Irvine, O.D. Dosser, J.B. Mullin: J. Cryst. Growth **68**, 245 (1984)
- 14.53 S. Fujita, S. Fujita: J. Cryst. Growth **145**, 552 (1994)
- 14.54 S. Fujita, A. Tababe, T. Sakamoto, M. Isemura, S. Fujita: J. Cryst. Growth **93**, 259 (1988)
- 14.55 S.J.C. Irvine, M.U. Ahmed, P. Prete: J. Electron. Mater. **27**, 763 (1988)
- 14.56 J. Wang, G. Du, B. Zhao, X. Yang, Y. Zhang, Y. Ma, D. Liu, Y. Chang, H. Wang, H. Yang, S. Yang: J. Cryst. Growth **255**, 293 (2003)
- 14.57 A.Y. Cho: J. Cryst. Growth **150**, 1 (1995)
- 14.58 C.T. Foxon: J. Cryst. Growth **251**, 1–8 (2003)
- 14.59 B.A. Joyce, T.B. Joyce: J. Cryst. Growth **264**, 605 (2004)
- 14.60 A.Y. Cho: *Molecular Beam Epitaxy* (AIP, New York 1994)
- 14.61 E.H.C. Parker: *The Technology and Physics of Molecular Beam Epitaxy* (Plenum, New York 1985)
- 14.62 B.A. Joyce, R.R. Bradley: Philos. Mag. **14**, 289–299 (1966)
- 14.63 J.R. Arthur: J. Appl. Phys. **39**, 4032 (1968)
- 14.64 A.Y. Cho: J. Vac. Sci. Technol. **8**, 31 (1971)
- 14.65 A.Y. Cho: Appl. Phys. Lett. **19**, 467 (1971)
- 14.66 J.W. Robinson, M. Illegems: Rev. Sci. Instrum. **49**, 205 (1978)
- 14.67 P.A. Barnes, A.Y. Cho: Appl. Phys. Lett. **33**, 651 (1978)
- 14.68 W.T. Tsang: Appl. Phys. Lett. **34**, 473 (1979)
- 14.69 A.Y. Cho, K.Y. Cheng: Appl. Phys. Lett. **38**, 360 (1981)
- 14.70 L.L. Chang, L. Esaki, W.E. Howard, R. Ludeke: J. Vac. Sci. Technol. **10**, 11 (1973)
- 14.71 H. Sakaki: J. Cryst. Growth **251**, 9 (2003)
- 14.72 A.Y. Cho: J. Appl. Phys. **41**, 2780 (1970)
- 14.73 M.D. Pashley, K.W. Habernern, J.M. Woodall: J. Vac. Sci. Technol. **6**, 1468 (1988)
- 14.74 J.J. Harris, B.A. Joyce, P.J. Dobson: Surf. Sci. **103**, L90 (1981)
- 14.75 J.H. Neave, B.A. Joyce, P.J. Dobson, N. Norton: Appl. Phys. **31**, 1 (1983)
- 14.76 C.T. Foxon, M.R. Boudry, B.A. Joyce: Surf. Sci. **44**, 69 (1974)
- 14.77 J.R. Arthur: Surf. Sci. **43**, 449 (1974)
- 14.78 C.T. Foxon, J.A. Harvey, B.A. Joyce: J. Phys. Chem. Solids **34**, 1693 (1973)
- 14.79 C.T. Foxon, B.A. Joyce: Surf. Sci. **50**, 434 (1975)
- 14.80 C.T. Foxon, B.A. Joyce: Surf. Sci. **64**, 293 (1977)
- 14.81 E.S. Tok, J.H. Neave, J. Zhang, B.A. Joyce, T.S. Jones: Surf. Sci. **374**, 397 (1997)
- 14.82 A.Y. Cho, J.R. Arthur: Prog. Solid State Chem. **10**(3), 157–191 (1975)
- 14.83 C.T. Foxon, B.A. Joyce: J. Cryst. Growth **44**, 75 (1978)
- 14.84 C.T. Foxon, B.A. Joyce, M.T. Norris: J. Cryst. Growth **49**, 132 (1980)
- 14.85 M.A. Herman, H. Sitter: *Molecular Beam Epitaxy*, Springer Ser. Mater. Sci., Vol.7 (Springer, Berlin, Heidelberg 1988) p.7
- 14.86 J. Saito, K. Nambu, T. Ishikawa, K. Kondo: J. Cryst. Growth **95**, 322 (1989)
- 14.87 M. Bafleur, A. Munoz–Yague, A. Rocher: J. Cryst. Growth **59**, 531 (1982)

- 14.88 Y. G. Chai, R. Chow: Appl. Phys. Lett. **38**, 796 (1981)
- 14.89 C. E. C. Wood, L. Rathburn, H. Ohmo, D. DeSimone: J. Cryst. Growth **51**, 299 (1981)
- 14.90 S. Izumi, N. Hayafuji, T. Sonoda, S. Takamiya, S. Mitsui: J. Cryst. Growth **150**, 7 (1995)
- 14.91 J. H. Neave, P. Blood, B. A. Joyce: Appl. Phys. Lett. **36**(4), 311 (1980)
- 14.92 C. R. Stanley, R. F. C. Farrow, P. W. Sullivan: *The Technology and Physics of Molecular Beam Epitaxy*, ed. by E. H. C. Parker (Plenum, New York 1985)
- 14.93 M. B. Panish: J. Electrochem. Soc. **127**, 2729 (1980)
- 14.94 A. R. Calawa: Appl. Phys. Lett. **38**(9), 701 (1981)
- 14.95 E. Veuhoff, W. Pletschen, P. Balk, H. Luth: J. Cryst. Growth **55**, 30 (1981)
- 14.96 N. Putz, E. Veuhoff, H. Heinicke, H. Luth, P. J. Balk: J. Vac. Sci. Technol. **3**(2), 671 (1985)
- 14.97 W. T. Tsang: Appl. Phys. Lett. **45**(11), 1234 (1984)
- 14.98 W. T. Tsang: J. Vac. Sci. Technol. B **3**(2), 666 (1985)
- 14.99 W. T. Tsang: Appl. Phys. Lett. **49**(3), 170 (1986)
- 14.100 T. H. Chiu, W. T. Tsang, J. E. Cunningham, A. Robertson: J. Appl. Phys. **62**(6), 2302 (1987)
- 14.101 W. T. Tsang, R. C. Miller: Appl. Phys. Lett. **48**(19), 1288 (1986)
- 14.102 J. S. Foord, C. L. Levoguer, G. J. Davies, P. J. Skevington: J. Cryst. Growth **136**, 109 (1994)
- 14.103 M. Weyers, J. Musolf, D. Marx, A. Kohl, P. Balk: J. Cryst. Growth **105**, 383–392 (1990)
- 14.104 R. J. Malik, R. N. Nottenberg, E. F. Schubert, J. F. Walker, R. W. Ryan: Appl. Phys. Lett. **53**, 2661 (1988)
- 14.105 F. Lelarge, J. J. Sanchez, F. Gaborit, J. L. Gentner: J. Cryst. Growth **251**, 130 (2003)
- 14.106 A. Y. Cho: J. Appl. Phys. **50**, 6143 (1979)
- 14.107 R. A. Stall, C. E. C. Wood, P. D. Kirchner, L. F. Eastman: Electron. Lett. **16**, 171 (1980)
- 14.108 R. Dingle, C. Weisbuch, H. L. Stormer, H. Morkoc, A. Y. Cho: Appl. Phys. Lett. **40**, 507 (1982)
- 14.109 G. B. Stringfellow, R. Stall, W. Koschel: Appl. Phys. Lett. **38**, 156 (1981)
- 14.110 C. R. Stanley, M. C. Holland, A. H. Kean, J. M. Chamberlain, R. T. Grimes, M. B. Stanaway: J. Cryst. Growth **111**, 14 (1991)
- 14.111 H. G. B. Hicks, D. F. Manley: Solid State Commun. **7**, 1463 (1969)
- 14.112 C. T. Foxon, J. J. Harris, D. Hilton, J. Hewett, C. Roberts: Semicond. Sci. Technol. **4**, 582 (1989)
- 14.113 K. Ploog: J. Cryst. Growth **81**, 304 (1987)
- 14.114 H. Tanaka, M. Mushiage: J. Cryst. Growth **111**, 1043 (1991)
- 14.115 J. Miller: III–Vs Rev. **4**(3), 44 (1991)
- 14.116 D. Bimberg, M. Grundmann, F. Heinrichsdorff, N. N. Ledentsov, V. M. Ustinov, A. R. Korsh, M. V. Maximov, Y. M. Shenyakov, B. V. Volovik, A. F. Tsatsalnikov, P. S. Kopiev, Zh. I. Alferov: Thin Solid Films **367**, 235 (2000)
- 14.117 S. Nakamura, T. Mukai, M. Senoh: Appl. Phys. Lett. **64**(13), 1689 (1994)
- 14.118 S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto: Jpn. J. Appl. Phys. **35**, 74 (1996)
- 14.119 H. Morkoc: J. Mater. Sci. Mater. El. **12**, 677 (2001)
- 14.120 S. E. Hooper, M. Kauer, V. Bousquet, K. Johnson, J. M. Barnes, J. Heffernan: Electron. Lett. **40**(1), 33 (2004)
- 14.121 N. Grandjean, M. Leroux, J. Massies, M. Laügt: Jpn. J. Appl. Phys. **38**, 618 (1999)
- 14.122 M. Kondow, K. Uomi, A. Niwa, T. Kitatani, S. Watahiki, Y. Yazawa: Jpn. J. Appl. Phys. **35**, 1273 (1996)
- 14.123 H. Riechert, A. Ramakrishnan, G. Steinle: Semicond. Sci. Technol. **17**, 892 (2002)
- 14.124 M. Kondow, T. Kitatani: Semicond. Sci. Technol. **17**, 746 (2002)
- 14.125 J. S. Harris, S. R. Bank, M. A. Wistey, H. B. Yuen: IEE Proc. Optoelectron. **151**(5), 407 (2004)
- 14.126 H. Ohno: J. Cryst. Growth **251**, 285 (2003)
- 14.127 H. J. Osten, E. Bugiel, O. Kirfel, M. Czernohorsky, A. Fissel: J. Cryst. Growth **278**, 18 (2005)
- 14.128 F.-J. Meyer zu Heringdorf, M. C. Reuter, R. M. Tromp: Nature **412**, 517 (2001)

15. Narrow-Bandgap II–VI Semiconductors: Growth

The field of narrow-bandgap II–VI semiconductors is dominated by the compound $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (CMT), although some Hg-based alternatives to this ternary have been suggested. The fact that CMT is still the preeminent infrared (IR) material stems, in part, from the fact that the material can be made to cover all IR regions of interest by varying the x value. In addition, the direct band transitions in this material result in large absorption coefficients, allowing quantum efficiencies to approach 100%. Long minority carrier lifetimes result in low thermal noise, allowing high-performance detectors to be made at the highest operating temperatures reported for infrared detectors of comparable wavelengths. This chapter covers the growth of CMT by various bulk growth techniques (used mainly for first-generation infrared detectors), by liquid phase epitaxy (used mainly for second-generation infrared detectors), and by metalorganic vapor phase and molecular beam epitaxies (used mainly for third-generation infrared detectors, including two-color and hyperspectral detectors). Growth on silicon substrates is also discussed.

15.1	Bulk Growth Techniques	304
15.1.1	Phase Equilibria	304
15.1.2	Crystal Growth	304
15.1.3	Material Characterization	306
15.2	Liquid-Phase Epitaxy (LPE)	308
15.2.1	Hg-Rich Growth	309
15.2.2	Te-Rich Growth	309
15.2.3	Material Characteristics	311
15.3	Metalorganic Vapor Phase Epitaxy (MOVPE)	312
15.3.1	Substrate Type and Orientation ...	315
15.3.2	Doping	316
15.3.3	In Situ Monitoring	317
15.4	Molecular Beam Epitaxy (MBE)	317
15.4.1	Double-Layer Heterojunction Structures	319
15.4.2	Multilayer Heterojunction Structures	319
15.4.3	CMT and CdZnTe Growth on Silicon	319
15.5	Alternatives to CMT	320
	References	321

The field of narrow-bandgap II–VI semiconductors is dominated by $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (CMT) (although some Hg-based alternatives to this ternary have been suggested and are discussed by Rogalski [15.1]). The reason that CMT is still the main infrared (IR) material is at least partially because this material can be made to cover all IR regions of interest by varying the value of x appropriately. This material also has direct band transitions that yield large absorption coefficients, allowing the quantum efficiency to approach 100% [15.2,3]. Furthermore, long minority carrier lifetimes result in low thermal noise allowing high-performance detectors to be made at the highest operating temperatures reported for infrared detectors of comparable wavelengths. These three major advantages all stem from the energy band structure of the material and they apply whatever device architecture is used. It has been shown that CMT is the third most stud-

ied semiconductor after silicon and gallium arsenide. This chapter covers the growth and characterization of CMT, mainly concentrating on the x region between 0.2 and 0.4 where the majority of applications are satisfied, and some of the Hg-based alternative ternary systems first described by Rogalski. The detectors made from these materials will be described in the chapter by Baker in this Handbook Chapt. 36.

In the first section the growth of CMT by various bulk techniques is reviewed. These include solid state recrystallization (SSR), Bridgman (plus ACRT, the accelerated crucible rotation technique), and the travelling heater method (THM). Despite the major advances made over the last three decades in the various epitaxial processes (liquid phase epitaxy, LPE, metalorganic vapor phase epitaxy, MOVPE, and molecular beam epitaxy, MBE), which are discussed in subsequent sections, bulk

techniques for CMT are still used for photoconductive devices and for large-size CdTe-based materials for CMT substrates (see Chapt. 16 on wide-bandgap II–VI compounds by Isshiki). CMT material grown by LPE is currently in production for large-area focal plane arrays (FPAs) for both terrestrial and space thermal imaging applications. More complex, fully doped heterostructures are being grown by MOVPE and MBE for both FPAs and novel nonequilibrium device structures. Both MOVPE

and MBE growth of device structures onto buffered silicon substrates is a topic of great interest currently, and this is briefly described.

For a fuller treatment of the topic matter in this chapter, plus further details of the characterization of CMT by various optical, electrical and structural means, the reader is referred to Chaps. 1 to 11 of [15.4], Sect. A of [15.5], Chaps. 1, 2, 10, 11, and 12 of [15.6], *Reine* [15.2], *Baker* [15.3], and Chapt. 7 of [15.7].

15.1 Bulk Growth Techniques

Despite the substantial progress that has taken place over the last three decades in various epitaxial growth techniques for CMT, bulk growth methods still play a key role when preparing device-quality material and producing substrates for these epitaxial techniques. A critical aspect of these techniques is the establishment of a thorough knowledge of the relevant phase equilibria, and a brief description of these phase diagrams for CMT is now given.

15.1.1 Phase Equilibria

The understanding of two types of equilibria is of critical importance. One is the solid compound in equilibrium with the gaseous phase (vapor growth) and the other is solid–liquid–gaseous equilibria (growth from liquid/melt). These phase equilibria also aid the understanding of post-growth heat treatments, either during cool-down to room temperature or during annealing processes used to adjust the stoichiometry, and hence the electrical properties of compounds. Solid–liquid–gaseous equilibria are described by three variables: temperature (T), pressure (p) and composition (x). It is, however, far easier to understand the interrelations between these parameters using two-dimensional (2-D) projections such as $T-x$, $p-T$ and $p-x$ plots.

A great deal of work has been carried out on the various phase equilibria ([15.5] contains reviews in this area). The liquidus and solidus lines in the pseudobinary HgTe–CdTe system are shown in Fig. 15.1. The wide separation between the liquidus and solidus, leading to marked segregation between the constituent binaries, CdTe and HgTe, was instrumental in the development of all the bulk growth techniques applied to this system. In addition to the solidus–liquidus separation, high Hg partial pressures are also influential during growth and post-growth heat treatments. A full appreciation of the

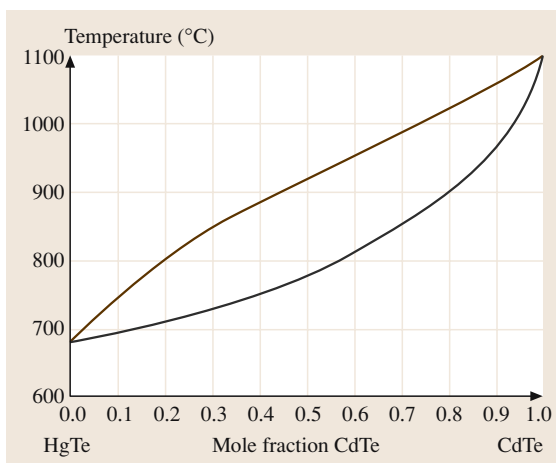


Fig. 15.1 $T-x$ phase diagram for the pseudo-binary CdTe–HgTe. (After [15.5])

$p_{\text{Hg}}-T$ diagram, shown in Fig. 15.2, is therefore essential. The curves are the partial pressures of Hg along boundaries for solid solutions of composition x where the solid solution is in equilibrium with another condensed phase as well as the vapor phase. For $x = 0.1$ and $10^3/T = 1.3 \text{ K}^{-1}$, CMT exists for Hg pressures between ≈ 0.1 (Te-saturated) and ≈ 7 (Hg-saturated) atm. Even at $x = 0.95$ and Te-saturated conditions, Hg is the predominant vapor species. It should also be noted that no solid solution contains exactly 0.5 atomic fraction Te. These features are highly significant for controlling the native defect concentrations and hence electrical properties in CMT.

15.1.2 Crystal Growth

Growth from the melt has constituted the vast majority of work on bulk growth of CMT. CMT grown by

several bulk methods is still in use, particularly for photoconductive detectors. Several historical reviews have been published [15.5, 8–10]. *Micklethwaite* [15.8] and *Kruse* [15.9] gave detailed discussions on the growth techniques and material characterization prior to 1980. Many techniques were tried, but three prime techniques survived these early attempts – **SSR**, Bridgman and **THM** – some of which are still in production, such as **SSR** in Germany and Bridgman in the UK. *Tennant et al.* [15.11] provided an authoritative view of the then current major issues in growth techniques. They concluded that the electrical performance had only recently been matched by **LPE** but pointed to the major drawbacks in bulk material of structural defects and size limitations for use in second-generation infrared detectors, such as focal plane arrays (**FPAs**).

Solid State Recrystallization (SSR)

This technique is used for ternary systems where there is a wide separation between solidus and liquidus, such as CMT. Other names that have been used for this process are quench anneal (**QA**) and cast recrystallize anneal (**CRA**). The term anneal is used in the first case to define a high-temperature grain-growth process, while in **CRA** it is a low-temperature process to adjust stoichiometry. In the basic technique, pure elements are

loaded into a cleaned silica ampoule and the charge is melted and rocked to ensure complete mixing. Charges are then normally quenched rapidly, into air or oil, to room temperature in the vertical orientation. This produces a dendritic structure that is reduced/removed by the recrystallization step that occurs at temperatures just below the melting point for many days. Grain growth occurs and microinhomogeneities in composition are removed. Care must be taken in the quenching stage to avoid pipes/voids that cannot be removed by the recrystallization step.

Alternatives to the basic **SSR** process have included ‘slush’ growth [15.12], high-pressure growth [15.13], incremental quenching [15.14] and horizontal casting [15.15]. In the ‘slush’ process, an initial homogenous charge is prepared and then held across the liquidus–solidus gap with the lower end solid and the upper end liquid. High-pressure growth (30 atm He gas) was used in an attempt to reduce structural defects. The ‘slush’ technique was used in production in the USA, while the incremental quenching technique has been used to provide large-diameter feed material for **THM** growth by *Colombo et al.* [15.16]. Other developments were made in the basic process (e.g. [15.17, 18]) but details available in the open literature are sketchy, due to proprietary constraints. *Tregilgas* [15.19] has given a detailed review of the **SSR** process.

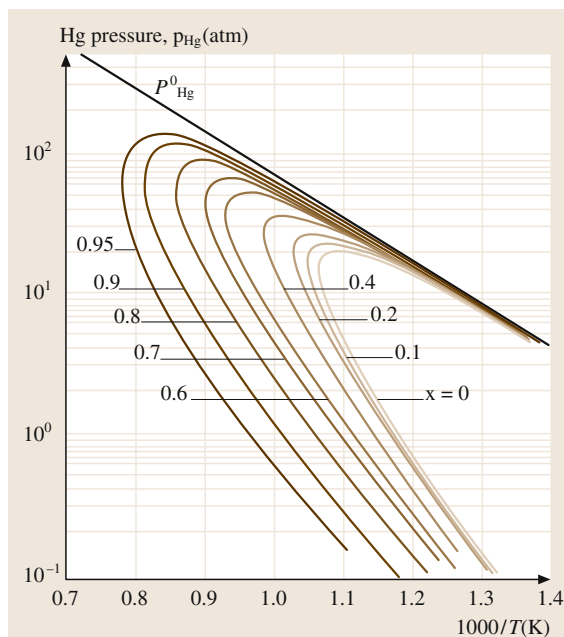


Fig. 15.2 p – T diagram for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (CMT). (After [15.5])

Bridgman (plus ACRT)

In the standard Bridgman process, elemental Cd, Hg and Te are loaded into a clean silica ampoule, homogenized by melting/rocking, and then frozen slowly from one end in a vertical system (Fig. 15.3, [15.20]) to produce a large-grained ingot [15.21]. There is marked segregation of CdTe with respect to HgTe (due to the wide gap between the solidus and liquidus) in the axial direction, but this leads to an advantage of the Bridgman process over other techniques: that material in both the main ranges of interest (3–5 and 8–12 μm : $x = 0.3$ and 0.2, respectively) is produced in a single run. The accelerated crucible rotation technique (**ACRT**) in which the melt is subjected to periodic acceleration/deceleration provided a means of stirring melts contained in sealed, pressurized ampoules. However, it has been recently suggested that **ACRT** does not automatically promote melt mixing [15.22]. They suggest that the balance of two time-dependent fluid dynamics phenomena decides whether mixing is promoted or not.

The first report of **ACRT** in CMT Bridgman growth was given in [15.23]. These effects were developed and reviewed in [15.5, 21]. Crystals were produced up to

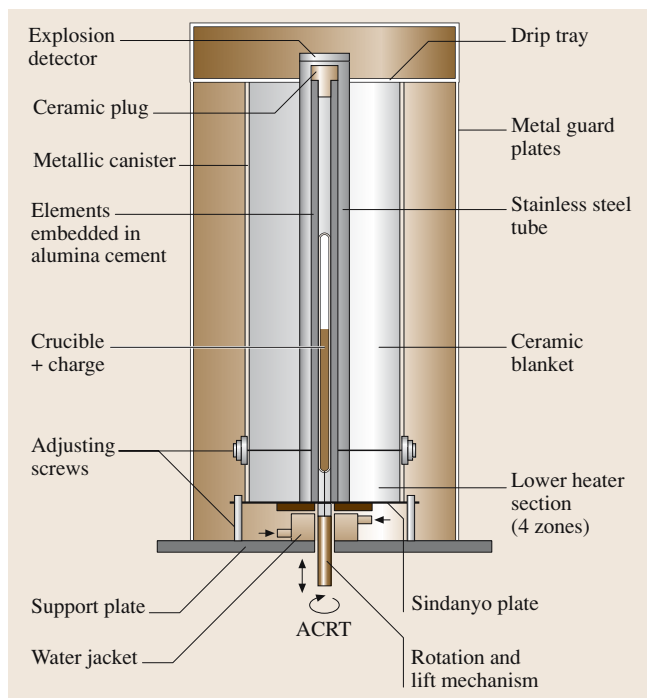


Fig. 15.3 Schematic of Bridgman growth kit for use with Hg-based compounds. (After [15.20])

20 mm in diameter and with x values of up to 0.6 in the tip regions of some crystals. Currently in this author's laboratory, crystals up to 21 mm in diameter and 200 mm in length weighing 0.5 kg are in routine production. Extensions to the basic ACRT process are enabling material of up to $x \approx 0.7$ to be produced in the first-to-freeze regions of crystals. These near-IR slices are highly transmissive and show a high degree of compositional uniformity ($\Delta x \approx 0.005$), see [15.24]. The use of ACRT produced material that enabled the fabrication of small two-dimensional arrays of photodiodes; this material was also used to demonstrate nonequilibrium detector operation and in the production of SPRITE detectors (Chapt. 11 of [15.6]).

Traveling Heater Method (THM)

Triboulet [15.25] (and his coworkers) developed this technique for CMT where diameters of up to 40 mm were accomplished and x values of up to 0.7 [15.26] for optical communication devices. Durand et al. [15.27] employed seeds to produce large oriented crystals, an advantage of this technique over other bulk methods. Gille et al. [15.28] adopted a slightly different approach. A pre-THM step was used to produce a quenched Te-rich

(53–60%) CMT melt. A first THM run then provided the source ingot, after removal of the tip and tail sections, for the final THM run. The entire growth procedure took several months but gave uniform material. This group also used rotation in the horizontal growth by THM with some success [15.29]. Two groups also applied ACRT to THM and obtained enhanced material properties [15.30,31].

15.1.3 Material Characterization

Compositional uniformity can be expressed in both the axial and the radial directions as $\Delta A(x)$ and $\Delta R(x)$, although the combination of the two clearly determines the yield of useful material. Assessment is normally by Fourier transform infrared (FTIR) spectrometry. The recrystallization process (in SSR) improves $\Delta A(x)$ from ± 0.05 to ± 0.02 [15.9] and removes the microinhomogeneity in x in both directions [15.32]. Independent control of axial and radial variations was claimed by Nelson et al. [15.33] to achieve ± 0.002 over 95% of slices up to 25 mm in diameter. Colombo et al. [15.14] quote $\Delta R(x)$ values of 0.005–0.01 in incremental-quenched material. Su et al. [15.34] report ± 0.008 for $\Delta A(x)$ and $\Delta R(x)$ in vertically cast material but better in horizontally cast material. Galazka [15.35] showed that reduced gravity gives smaller $\Delta R(x)$, and they also noted that constitutional supercooling at the first-to-freeze end is also reduced and the growth is diffusion-limited. The 'slush' process [15.8] produces $\Delta R(x)$ of ± 0.005 but only over the region of axially uniform material ≈ 3 mm in length.

Bartlett et al. [15.36] grew SSR and Bridgman crystals over a wide range of growth rates and concluded that $\Delta R(x)$ was due to the combination of a concave (to the melt) growth interface and density-driven convective flow within a boundary layer close to the interface. This, added to the normal segregation of the low melting point HgTe to the center of the concave interface, led to the observed variation. It was also concluded that crystals with simultaneous small $\Delta A(x)$ and $\Delta R(x)$ were not possible; slow growth gave low $\Delta R(x)$ while fast growth minimized $\Delta A(x)$. In Bridgman material grown with ACRT, an 80 mm length at $x = 0.21$ with both $\Delta A(x)$ and $\Delta R(x)$ within ± 0.002 was reported by Capper [15.21]. This was linked to a flat interface produced by strong Ekman flow in the interface region, although all Bridgman crystals have concave interfaces next to the ampoule walls [15.21]. A 20 mm-diameter ACRT crystal was grown that exhibited $\Delta R(x) = \pm 0.003$ at $x = 0.22$ [15.21].

Triboulet et al. [15.10] quoted values of ± 0.02 (over 3 cm at $x = 0.2$) and ± 0.002 for $\Delta A(x)$ and $\Delta R(x)$, respectively, for **THM** growth at 0.1 mm/h. Corresponding figures from *Colombo et al.* [15.16] for incremental-quenched starting material and a similar growth rate of 2 mm/d were ± 0.01 and ± 0.005 . For the very slow grown material produced by *Gille et al.* [15.28], variations in composition were within experimental error for $\Delta R(x)$ while $\Delta A(x)$ was ± 0.005 . *Royer et al.* [15.30] obtained improved radial uniformity with the addition of **ACRT** to **THM**, but in the work of *Bloedner and Gille* [15.31] both $\Delta A(x)$ and $\Delta R(x)$ were only as good as non-**ACRT** crystals.

With regard to electrical properties, as-grown materials from **SSR**, **ACRT** Bridgman (with $x < 0.3$) and **THM** are highly p-type in nature, believed to be due to metal vacancies. These types of materials can all be annealed at low temperatures, in the presence of Hg, to low n-type levels, indicating that the p-type character is due to metal vacancies. By contrast, Bridgman material is n-type as-grown [15.21] and the residual impurity donor level is found to be $< 5 \times 10^{14} \text{ cm}^{-3}$. *Higgins et al.* [15.18] have shown that n-type carrier concentrations can be $< 10^{14} \text{ cm}^{-3}$ at $x \approx 0.2$ in melt-grown material. A great deal of work has been done on the annealing behavior of **SSR** material, see *Tregilgas* [15.19]. n-type levels of $2 \times 10^{14} \text{ cm}^{-3}$ after Hg annealing $x = 0.2$ and higher- x material were reported by *Nguyen Duy et al.* [15.26], while *Colombo et al.* [15.16] quote $4 \times 10^{14} \text{ cm}^{-3}$. *Durand et al.* [15.37] note that **THM** growth at 600 °C results in p-type behavior, but growth at 700 °C gives n-type material. In this author's laboratory, current 20 mm-diameter **ACRT** material has reached mid- 10^{13} cm^{-3} levels with high mobility after a normal low-temperature Hg anneal step. This is thought to reflect the improvements made in the purity of starting elements over the recent years. The basic p to n conversion process [15.19] has recently been extended by *Capper et al.* [15.38] to produce analytical expressions that can account for the temperature, donor level and composition dependencies of the junction depth.

High minority n-type carrier lifetimes have been reported in all three types of material. *Kinch* [15.39] noted values $> 1 \mu\text{s}$ (at 77 K) in $x = 0.2$ **SSR** crystals, while *Triboulet et al.* [15.10] quote $3 \mu\text{s}$ in **THM** material. *Pratt and coworkers* [15.40,41] found high lifetimes, up to $8 \mu\text{s}$ in $x = 0.23$ Bridgman and **ACRT** material with n-type levels of $1\text{--}6 \times 10^{14} \text{ cm}^{-3}$, and up to $30 \mu\text{s}$ for equivalent $x = 0.3$ material (192 K).

In terms of extrinsic doping, most elements are electrically active in accordance with their position in the

Periodic Table. This is true in **SSR** material for Group V and VII elements only after a high-temperature treatment, and is linked to the stoichiometry level at the growth temperature (in other words, those elements that substitute on Te lattice sites have to be forced onto the correct sites in Te-rich material). Group I and III elements are acceptors and donors, respectively, on the metal sites. There is evidence that some Group I elements can migrate at low temperatures to grain boundaries or to the surface of samples [15.21]. In **ACRT** crystals, Groups I and III are acceptors and donors, respectively, on the metal sites, as they are in Bridgman, with the exception of Au [15.21]. Groups V and VII are inactive dopants in those portions of **ACRT** crystals that are Te-rich as-grown ($x < 0.3$), but are active dopants for $x > 0.3$ where metal-rich conditions prevail, as found in Bridgman material. In doped material, grown by either standard Bridgman or **ACRT**, acceptor ionization energies were found to be lower than undoped counterparts [15.21]. Segregation of impurities in **SSR** is very limited due to the initial fast quench step. By contrast, Bridgman and **THM** benefit from marked segregation of impurities due to their relatively slow growth rate. Impurity segregation behavior was affected by **ACRT** [15.21], in general, segregation coefficients decrease in **ACRT** crystals, when compared to standard Bridgman. This segregation leads to very low levels of impurities in both **THM** [15.25] and Bridgman/**ACRT** material [15.42].

Vere [15.43] has reviewed structural properties and noted grains, subgrains, dislocations, Te precipitates and impurities on dislocations as major problems. Grain boundaries act as recombination centers and generate noise and dark current in devices. Subgrain sizes cover 50–500 μm in both **SSR** and Bridgman material. *Williams and Vere* [15.44] showed how a recrystallization step at high pressure and temperatures of $> 600^\circ\text{C}$ coalesces subgrains and eventually eliminates them. Tellurium precipitation was extensively studied [15.44,45] and was summarized in [15.44]. Precipitates were found to nucleate on dislocations during the quench from the recrystallization step but a 300 °C anneal dissolves the precipitates, generating dislocation loops that climb and lead to dislocation multiplication.

Quenching studies in Bridgman/**ACRT** crystals grown in flat-based ampoules [15.21] revealed not only a flat growth interface but also that the slow-grown material produced prior to quenching was single crystal. This demonstrated the power of Ekman stirring and the importance of initiating the growth of a single crystal grain. In Bridgman material, which is close

to stoichiometric (Te excess < 1%), little Te precipitation occurs. Above this limit, precipitates/inclusions of 0.1–0.5 mm in size are seen. Recent work in this author's laboratory [15.24] has demonstrated flat solid–liquid interfaces in the $x \approx 0.21$ region of 20 mm-diameter ACRT crystals, mirroring the high level of compositional uniformity in these crystals.

Triboulet et al. [15.10] found no Te precipitates in their THM material, despite using a Te zone. Several groups [15.14, 46] noted an increase in dislocation density and in rocking curve widths on moving from the center to the edge of THM crystals. The addition of ACRT to the THM process led to increased growth rates with no detriment to the Te precipitation or grain structure.

We have recently [15.24, 47] extended our ACRT growth process to produce high- x material for various near-IR applications. This necessitated improving the ampoule sealing procedure to manage the higher pressures caused by the higher growth temperatures. Assessment techniques were extended beyond the normal FTIR wavelength characterization to include imaging of defects using a mid-IR camera system.

To summarize, the two main problems with all bulk material are the limitation on size – only THM has reproducibly produced material above 20 mm in diameter – and the fact that it is rarely completely single-crystal in nature. These two limitations led to the drive in the late 1970s and early 1980s to develop various epitaxial growth techniques.

15.2 Liquid-Phase Epitaxy (LPE)

The process of LPE has emerged as the predominant materials growth technology for the fabrication of second-generation CMT focal plane arrays (FPAs). Material can now be routinely grown for high-performance photoconductive, photovoltaic, and laser-detector devices over the entire 2–18 μm spectral region. This is manifested by increased size (over bulk material), enabling large-array formats to be made; reduced cycle time; better material characteristics in terms of composition uniformity, crystal quality, and electrical properties, making the material no longer the limiting factor in FPAs; and the realization of some advanced device structures, either to simplify the FPA design (back-side-illuminated detectors) or to enhance detector performance (double-layer heterostructures).

Two completely different technical approaches have been pursued: growth from Te solutions and growth from Hg solutions. One major advantage of the Hg-solution

technology is its ability to produce layers of excellent surface morphology due to the ease of melt decanting. Two additional unique characteristics, essential for the fabrication of high-performance double-layer heterojunction (DLHJ) detectors by LPE, are low liquidus temperature (< 400 °C), which makes growth of the cap layer feasible, and ease of incorporating both p-type and n-type temperature-stable impurity dopants, such as As, Sb and In, during growth. While layers grown from Hg-rich solutions are easily doped with Group VB elements with high solubility [15.48], layers grown from tellurium-rich solutions are not. Group IIIB elements, in particular indium, are easily incorporated from both solutions.

A number of LPE approaches have been used to grow both thin and thick films. The principal LPE techniques used are tipping, dipping and sliding boat techniques. A schematic diagram of these three processes is shown in Fig. 15.4. The tipping and dipping techniques have been

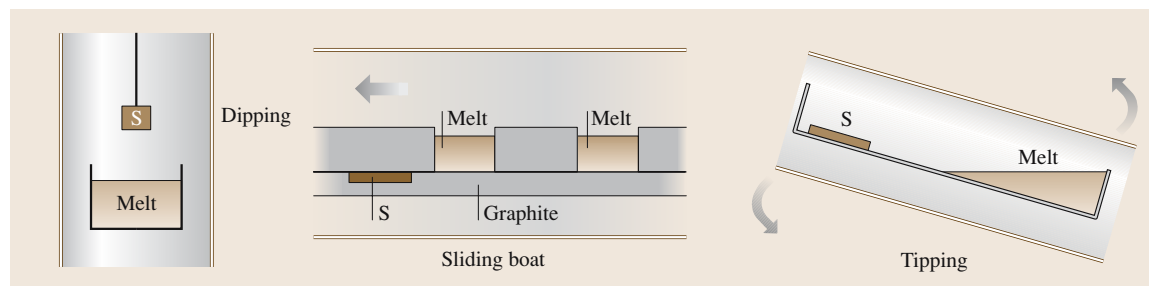


Fig. 15.4 Dipping, sliding boat and tipping LPE arrangements

implemented by using both tellurium- and mercury-rich solutions, whereas only tellurium-rich solutions have been used with the sliding boat. Both dipping and sliding boat Te-rich techniques are still in widespread use.

Extensive experimental phase diagram and thermodynamic data have been critically reviewed, along with the results calculated by the associated solution model [15.49]. As in bulk growth, full knowledge of the solid–liquid phase relation is essential for proper use of solution-growth processes. In addition, the solid–vapor and liquid–vapor phase relations are of practical importance, especially in view of the high Hg pressure in the growth process and the effect of the vapor of constituent components upon post-growth annealing and the consequent electrical properties. *Astles* [15.50] reviewed the experimental data on Te-rich LPE growth at 460 to 550 °C.

15.2.1 Hg-Rich Growth

For CMT growth from Hg-rich melts, the design and operation of a system is dominated by the consideration of the high vapor pressure of Hg, which comprises $\geq 90\%$ of the growth solution. A secondary but related factor is the requirement to minimize melt composition variation during and between growths due to solvent or solute loss. These factors led to the evolution of a vertical high-pressure furnace design with a cooled reflux region. The furnace has to provide a controllable, uniform and stable thermal source for the melt vessel, which has to be capable of maintaining at least 550 °C continuously. The cylindrical melt vessel consists of a high-strength stainless steel chamber lined with quartz. Such systems are capable of containing about 10–20 kg of melt at 550 °C for several years with no degradation in melt integrity or purity.

The system must be pressurized and leak-free to keep the Hg-rich melt from boiling or oxidizing at temperatures above 360 °C. Typical pressures range up to 200 psi and the pressurization gas may be high-purity H_2 or a less explosive reducing gas mixture containing H_2 . The melts are always kept saturated and are maintained near to the growth temperature and pressure between successive runs. The prepared substrates are introduced into the melt through a transfer chamber or air lock. A high-purity graphite paddle with externally actuated shutters holds the substrates. The paddle with the shutters closed is not gas-tight but protects the substrates from undue exposure to Hg vapor/droplets, or other condensing melt components. The paddle assembly can be lowered into the melt and rotated to stir it.

In normal operation, the high-purity melt components are introduced into the clean melt vessel at room temperature and the system is sealed, evacuated and pressurized. The temperature of the furnace is raised above the predicted melting point and held constant until all of the solute dissolves. The amount of material removed from the melt during each growth run is relatively insignificant. Optimum layer smoothness occurs on polished lattice-matched CdZnTe substrates oriented close to the $\langle 111 \rangle$ plane. Growth begins by lowering the paddle plus substrates into the melt and allowing thermal equilibrium to be reached while stirring. A programmed ramp then reduces the melt temperature to the required level, at which point the shutters are opened and the substrates are exposed to the melt. The growth rate and layer thickness are determined mainly by the exposure temperature relative to the saturation point and the total growth range. The composition of the layer and its variation are determined mainly by the melt composition and its thermal uniformity. Large melts allow the production of layer areas of up to 30 cm² with excellent compositional and thickness uniformity, and allow dopant impurities to be accurately weighed for incorporation into layers and to maintain stable electrical characteristics over a long period of time. Four layers with a total area of 120 cm² can be grown in a single run [15.51]. *Norton et al.* [15.52] have also scaled-up for the growth of cap layers from Hg-rich solutions, with each reactor capable of growth on four 24 cm² base layers per run.

15.2.2 Te-Rich Growth

A number of problems encountered with bulk crystal growth techniques are solved using CMT growth from tellurium-rich solution. The most important of these is the reduction of the Hg vapor pressure over the liquid by almost three orders of magnitude at the growth temperature. Growth from Te-rich solutions is used in three embodiments: dipping, tipping and sliding boat technologies (Fig. 15.4). While the tipping process may be used for low-cost approaches, it is not as widely used as the sliding boat and the dipping techniques. A comparison of the three techniques is shown in Table 15.1.

Current dipping reactors are capable of growing in excess of 60 cm² per growth run and are kept at temperature for long periods, > 6 months. Melts, on the other hand, last a very long time, > 5 years. A sensor-based reactor capable of growing CMT thick layers at relatively high production volumes and with excellent

Table 15.1 Comparison of three LPE growth techniques

LPE growth technique	Advantages	Disadvantages
Sliding boat	Thickness control Large area Double layers	Thick layers difficult to grow Substrate thickness/planarity control High cost
Tipping	Thickness control Lowest investment for experimental investigation Closed system/good p_{Hg} control	High cost Scale up
Dipping	Thick layer growth High throughput Low cost Large/flexible area	Thin-layer thickness control Double layers

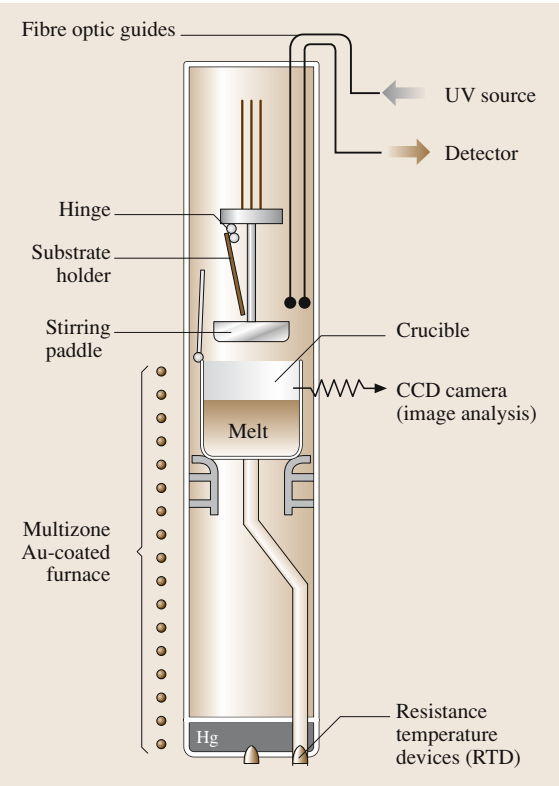


Fig. 15.5 Schematic diagram of a dipping LPE reactor showing the Te-rich melt, the mercury reservoir and positions of the sensors

reproducibility was described by Westphal et al. [15.53], Fig. 15.5. The reactors utilize resistance temperature devices (RTDs) for temperature control, a UV-vis spec-

trometer to measure the mercury partial pressure, and a video camera to aid in the observation of the liquidus temperature. The RTDs were used to control the temperature within a maximum range of $\pm 0.005^\circ\text{C}$ at a temperature of $\approx 480^\circ\text{C}$. The melts are about 4 kg with a liquidus temperature of $\approx 480^\circ\text{C}$ for a solid CMT composition of $x = 0.225$. Accurate mercury vapor pressure control ensures constant liquidus and thus constant liquid composition.

Layers produced from Te-rich solutions are usually grown by isothermal supersaturation, programmed cooling techniques, or a combination of both. The details of the mass transport of Hg-Cd-Te solutions during LPE growth were presented by Shaw [15.54]. He predicted that isothermal growth of thin layers from supersaturated solutions would yield layers with uniform composition but that the growth rate would change during growth. While the growth rate is more uniform for layers grown by the programmed cooling technique, the composition gradient is higher.

The dipping growth process comprises: 1) mixing, 2) liquidus measurement, 3) meltback, and 4) growth. Because of the relatively large melt, stirring is required to mix the solution. The choice of growth method depends on the desired properties of the films. Typically, isothermal supersaturation is preferred when growing thin layers, whereas programmed cooling or a combination of both is used for thicker films. Colombo et al. [15.55] used a melt-tracking technique to control the composition of the solid.

Sliding-boat LPE normally consists of a graphite boat with a recess in the base to hold the substrate wafer and a movable block with wells that contain the

LPE solution and that allows the solution to be brought into contact with the substrate and then wiped off after growth. The main advantages of the technique are the efficient use of solution and the possibility of growing multilayers. The main disadvantages are the need for careful machining of the boat components in order to obtain efficient removal of the solution after growth and the need for precisely sized substrate wafers to fit into the recess in the boat. The sliding boat growth process has several variants, but essentially a polished substrate is placed into the well of a graphite slider and the Te-rich solution is placed into a well in the body of the graphite boat above the substrate and displaced horizontally from it. Normally, a separate well contains the HgTe charge to provide the Hg vapor pressure needed during growth and during cool-down to control the stoichiometry. The boat is then loaded into a silica tube that can be flushed with nitrogen/argon prior to the introduction of H_2 for the growth phase. The furnace surrounding the work tube is slid over the boat, and the temperature is increased to 10–20°C above the relevant liquidus. At that point, a slow temperature ramp (2–3°C/h) is initiated, and when the temperature is close to the liquidus of the melt the substrate is slid under the melt and growth commences. After the required thickness of CMT has been deposited (typical growth rates are 9–10 $\mu\text{m/h}$), the substrate is withdrawn and the temperature is decreased to an annealing temperature (to fix the p-type level in the as-grown material) before being reduced rapidly to room temperature. Layer thicknesses of 25–30 μm are normally produced for loop-hole diode applications [15.56].

15.2.3 Material Characteristics

Good composition uniformity, both laterally and in depth, is essential in order to obtain the required uniform device performance. Growth parameters that need to be optimized in Hg-rich LPE include the degree of supercooling and mixing of the melt, the geometrical configuration of the growth system, the melt size and the phase diagram. The standard deviation of the cut-off wavelength, for 12-spot measurements by Fourier transform infrared (FTIR) transmission at 80 K across a 30 cm^2 LWIR layer, is reported as 0.047.

Composition control and the uniformity of layers grown by dipping Te-rich LPE is one of the strengths of this process. The cut-off wavelength reproducibility is typically $10.05 \pm 0.18 \mu\text{m}$. Dipping Te-rich LPE is mainly used to grow thick films, about 100 μm , hence thickness control is not one of its advantages. Thickness

control is about $\pm 15\%$ for layers of $< 20 \mu\text{m}$ due to the relatively large amount of solidified material.

For Te-rich sliding-boat LPE, layers of $\approx 30 \mu\text{m}$ thickness can show wavelength uniformity at room temperature of $6.5 \pm 0.05 \mu\text{m}$ over 90% of the area of $20 \times 30 \text{ mm}$ layers [15.57].

The ease of decanting the Hg-rich melt after layer growth results in smooth and specular surface morphology if a precisely oriented, lattice-matched CdZnTe substrate is used. Epitaxial growth reproduces the substructure of the substrate, especially in the case of homoepitaxy.

The dislocation density of LPE CMT and its effects on device characteristics have received much attention [15.58, 59]. The dislocation density is dominated mainly by the dislocations of the underlying substrate [15.60]. For layers grown on substrates with $\text{ZnTe} \approx 3-4\%$, the dislocations are present only in the interface region and the dislocation density is close to that of the substrate. For layers grown on substrates with $\text{ZnTe} > 4.25\%$, dislocation generation is observed within a region of high lattice parameter gradient. The same variation of dislocation density with depth is seen for sliding-boat LPE material [15.56] (typical values are $3-7 \times 10^4 \text{ cm}^{-2}$).

For the production of heterostructure detectors with CMT epitaxial layers, it is essential that proper impurity dopants be incorporated during growth to form well-behaved and stable p–n junctions. An ideal impurity dopant should have low vapor pressure, low diffusivity, and a small impurity ionization energy. Group V and Group III dopants – As and Sb for p-type and In for n-type – are the dopants of choice. Hg-rich melts can be readily doped to produce n- and p-type layers; the solubilities of most of the useful dopants are significantly higher than in Te-rich solutions, most notably for Group V dopants, which are among the most difficult to incorporate into CMT.

Accurate determinations of dopant concentration in the solid involve the use of Hall effect measurements and secondary ion mass spectrometry (SIMS) concentration profiles. Measurements on the same sample by the two techniques are required to unequivocally substantiate the electrical activity of impurity dopants. The ease of incorporating Group I and Group III dopants into CMT, irrespective of non-stoichiometry, has been confirmed experimentally [15.5, 61].

The excess carrier lifetime is one of the most important material characteristics of CMT since it governs the device performance and frequency response. The objective is to routinely produce material with a lifetime that

is limited by Auger processes, or by the radiative process in the case of the medium-wavelength infrared (MWIR) and short-wavelength infrared (SWIR) material [15.62]. It has been reported that intentionally impurity-doped LPE CMT material grown from As-doped Hg-rich melts can be obtained with relatively high minority carrier lifetimes [15.63]. The 77 K lifetimes of As-doped MWIR ($x = 0.3$) CMT layers are significantly higher than those of undoped bulk CMT and are within a factor of two of theoretical radiative lifetimes. Various annealing schedules have been proposed recently [15.38] that may lead to a reduction in Shockley–Read traps with a consequent increase in lifetime, even in undoped material. Lifetimes in In-doped MWIR CMT were also found to exhibit an inverse linear dependence on the doping concentration [15.63], with $N_d\tau$ products similar to $N_a\tau$ products observed for the As-doped material. The lifetimes of LWIR In-doped LPE material are typically limited by the Auger process at doping levels above 10^{15} cm^{-3} .

The first heterojunction detectors were formed in material grown by Hg-rich LPE [15.64]. For double-layer heterojunctions (DLHJ), a second LPE (cap) layer is grown over the first (base) layer. With dopant types and layer composition controlled by the LPE growth process, this approach offers great flexibility (p-on-n or n-on-p) in junction type and in utilizing heterojunction formation between the cap and absorbing base layers to optimize detector performance. The key step in the process is to grow the cap layer doped with slow-diffusing impurities,

In for an n-type cap layer, and As or Sb for a p-type cap layer.

For future large-area FPAs, Si-based substrates are being developed as a replacement for bulk CdZnTe substrates. This effort is directed at improvements in substrate size, strength, cost and reliability of hybrid FPAs, particularly during temperature cycling. These alternative substrates, which consist of epitaxial layers of CdZnTe or CdTe on GaAs/Si wafers [15.65] or directly onto Si wafers [15.66], are particularly advantageous for the production of large arrays. High-quality epitaxial CMT has been successfully grown on the Si-based substrates by the Hg-melt LPE technology for the fabrication of p-on-n DLHJ detectors. The first high-performance 128×128 MWIR and LWIR arrays were demonstrated by Johnson et al. [15.67]. MWIR arrays as large as 512×512 and 1024×1024 have also been produced [15.68]. A bias-selectable two-color (LWIR/MWIR) detector structure was first fabricated by growing three LPE layers from Hg-rich melts in sequence on a bulk CdZnTe substrate [15.68]. The structure forms an n-p-n triple-layer graded heterojunction (TLHJ) with two p-n junctions, one for each spectral band (color).

Destefanis et al. [15.69] have recently described their work on large-area and long linear FPAs based on Te-rich sliding-boat LPE material. Using a $15 \mu\text{m}$ pitch they were able to produce 1000×1000 MW arrays of photodiodes and 1500×2 MW and LW long linear arrays.

15.3 Metalorganic Vapor Phase Epitaxy (MOVPE)

Metalorganic vapor phase epitaxy (MOVPE) of CMT is dominated by the relatively high vapor pressures of mercury that are needed to maintain equilibrium over the growing film. This arises from the instability of HgTe compared with CdTe, and requires much lower growth temperatures than are usual for more stable compounds. The MOVPE process was developed as a vapor phase method that would provide sufficient control over growth parameters at temperatures below 400°C , the main advantage being that the elements (although not mercury) can be transported at room temperature as volatile organometallics and react in the hot gas stream above the substrate or catalytically on the substrate surface. The first mercury chalcogenide growth by MOVPE [15.70,71] was followed by intense research activity that has brought the technology to its current state of maturity.

Although the MOVPE reaction cell conditions are far from equilibrium, an appreciation of the vapor–solid equilibrium can determine the minimum conditions needed for growth. This is particularly important with the mercury chalcogenides, where the relatively weak bonding of mercury causes a higher equilibrium vapor pressure. The equilibrium pressures of the component elements are linked and there is a range of pressures over which the solid remains in equilibrium as a single phase. At MOVPE growth temperatures, the pressure can vary by three orders of magnitude and remain in equilibrium with a single phase of HgTe. However, the Te_2 partial pressure varies across the phase field in the opposite sense to the Hg partial pressure.

The maintenance of vapor pressure equilibrium within the reactor cell does not automatically lead to the growth of an epilayer, but it does enable us to elimi-

nate conditions that would not be suitable. For example, the supply of Hg to the substrate for growth at 400 °C must be greater than the minimum equilibrium pressure of 10^{-2} bar. This pressure is higher than would be conveniently delivered from a mercury organometallic source that would typically be less than 10^{-3} bar. All narrow-gap II–VI epilayers, except for a few examples of very low temperature growth, have used a heated mercury source that is capable of delivering the required partial pressure.

The criteria for precursors that are generally applicable to MOVPE are volatility, pyrolysis temperature, stability and volatility of organic products and purity. The latter depends on synthesis and purification routes for the precursors, but techniques such as adduct purification of the Group II organometallics have made a major contribution to epilayer purity. A good review of this topic is given by Jones [15.72].

The precise control over layer properties that is required for uniform detector arrays and tuning of the response wavelength band can only be achieved if the cadmium concentration is controlled to better than 1%. Improvements in reactor design, growth techniques (such as IMP – the interdiffused multi-layer process), choice of precursors and operating conditions have all contributed to successful targeting of the composition and thickness. However, these conditions are difficult to achieve on a run-to-run basis due to drifts in calibration, and they will require additional levels of control. Recent developments with in situ monitoring

have been instrumental to gaining a better understanding of this technology and to reducing variance in epilayer properties. This has been achieved with better system monitoring to measure organometallic concentration and epilayer monitoring with laser reflectometry or ellipsometry [15.73].

A typical example of a horizontal reactor cell with an internal mercury source is shown in Fig. 15.6. The reactor cell wall temperature is also determined by the thermodynamics, which requires a minimum temperature in order to avoid condensation of mercury, and these requirements can be simply determined from the phase diagram. The organometallic supply shows the DMCD injected beyond the Hg source, which is the most common practice used to avoid reaction in the Hg zone.

Much of the work on precursors has been concerned with the thermal stability of the tellurium source, as this limits the growth temperature. From thermodynamic considerations, it would be desirable to grow at low temperature (below 400 °C) where the requirement on Hg equilibrium pressure will be lower. The initial demonstration of growth using MOVPE was with diethyl telluride (DET_e), which requires a temperature above 400 °C for efficient pyrolysis [15.71].

Despite some high hopes of reducing the growth temperature to 200 °C, where the required equilibrium mercury vapor pressure would be as low as 10^{-7} bar, the most successful tellurium source has been diisopropyltelluride (DIPT_e), which is used for growth between 350 and 400 °C. In practice, the DiPT_e bubbler is held

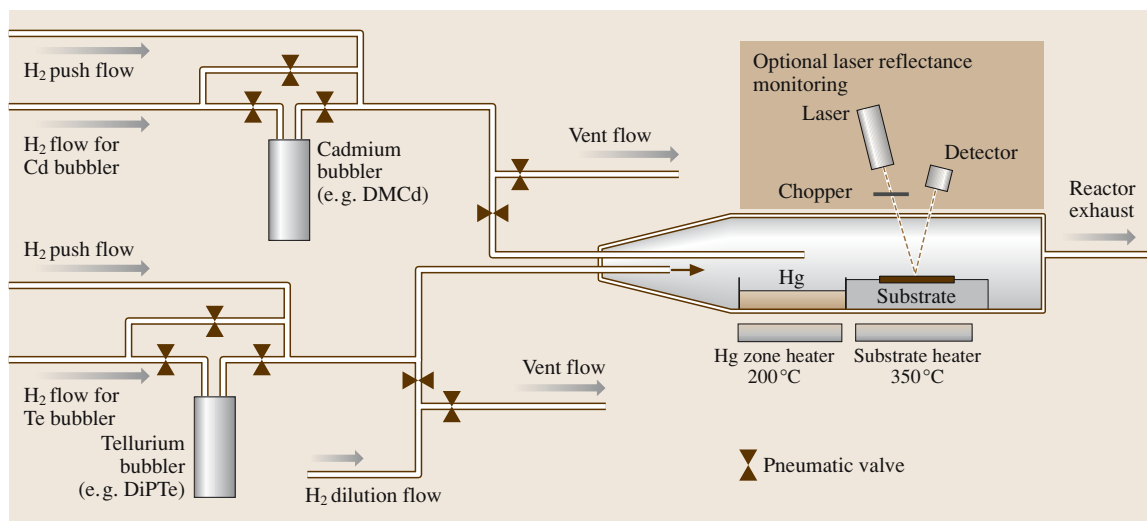


Fig. 15.6 Horizontal reactor cell with internal mercury source and gas supply suitable for the growth of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloy. (After S.J.C. Irvine, in [15.4])

at a temperature above ambient (up to 30°C) and the lines to the reactor are heated to avoid condensation of the precursor in the feed lines. Care must be taken, however, not to overheat the lines and cause premature pyrolysis.

Although most of the development of precursors in narrow gap II–VI semiconductors has been concerned with the tellurium source, there have been some important developments with the Group II sources. The cadmium source dimethyl cadmium (**DMCd**) appears to work well down to 250°C, although this probably relies on reaction chemistry with the tellurium source.

Unlike **MOVPE** growth of other II–VI and III–V semiconductors, the growth of an alloy in narrow-gap II–VIs can be achieved by one of two alternative processes, direct alloy growth (**DAG**) and the interdiffused multilayer process (**IMP**). The requirement on alloy uniformity is critical, because this will determine the cut-on wavelength of a detector, and for focal plane arrays there needs to be uniformity of cut-on wavelength across the array. The **DAG** approach suffers from disadvantages in the thermodynamics and reaction chemistry in that the reaction rate of the tellurium precursor is very different when growing the mercury-containing binary compared with Zn, Cd, or Mn. This leads to a preferential depletion of the non-Hg Group II species as the gas flow passes over the substrate, and consequent compositional changes in the flow direction.

The interdiffused multilayer process [15.75] separates the growth of HgTe and CdTe so that the growth of the binaries can be independently optimized. The procedure entails the growth of alternate layers of HgTe and CdTe that interdiffuse during the time of growth to give a homogeneous epitaxial layer. The **IMP** approach relies on the very high interdiffusion coefficients in the CMT pseudobinary [15.38], and the basic principles are as follows:

1. Interdiffusion must occur at the growth temperature, and a nominal time at the end of growth for the last periods to interdiffuse must be allowed.
2. There should be no residual compositional or structural modulation attributable to the **IMP** oscillations.
3. Interdiffusion at heterointerfaces should be no greater than for direct alloy growth (**DAG**) at that temperature.
4. Flow velocity is adjusted between the two binary growths in order to optimize growth uniformity.

The alloy composition is fixed by adjusting the times for growth of the HgTe and CdTe **IMP** layers. Figure 15.7 shows the uniformity of wavelength over a 3-inch-diameter **MOVPE** layer grown on a GaAs:Si substrate [15.74]. 2-D contour maps of the cut-off wavelength of 3–5 μm wavelength slices (contour steps = 0.1 μm) showed that without rotation

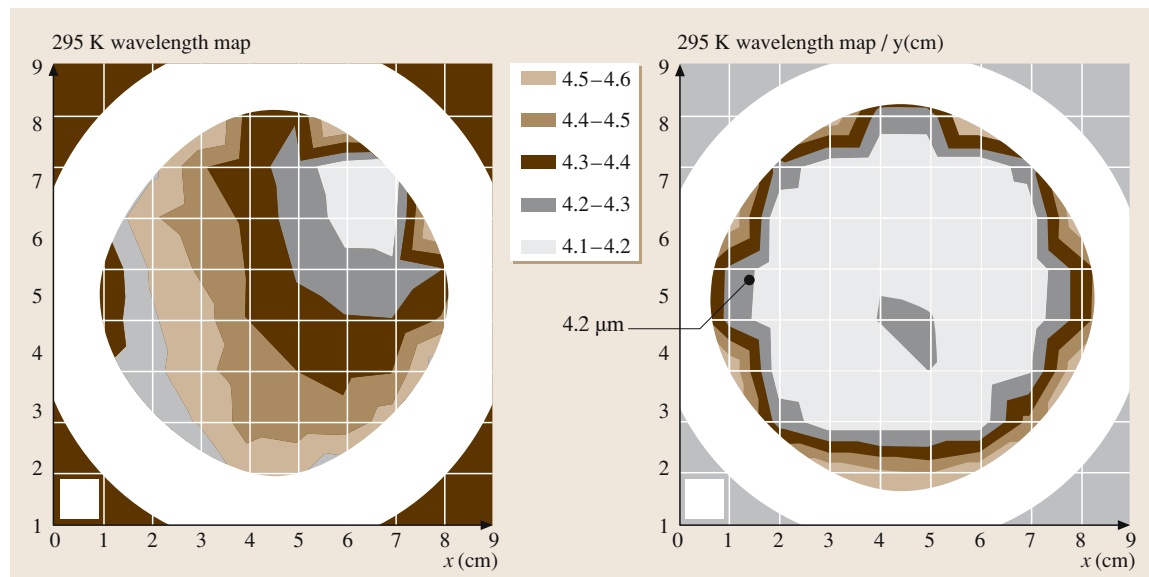


Fig. 15.7 Uniformity of wavelength over a three-inch-diameter **MOVPE** layer grown on a GaAs:Si substrate. (After [15.74])

(Fig. 15.7a) there is a gradual decrease in composition from upstream to downstream such that the compositional uniformity was within ± 0.022 in x . The thickness uniformity was typically $\pm 5-6\%$. However, by introducing substrate rotation, the uniformity improved dramatically such that, to within a few millimeters of the wafer edge, the composition was uniform to within ± 0.004 (Fig. 15.7b). The thickness uniformity also improved to within $\pm 2-3\%$. This was compatible with the production of twelve sites of large-format 2-D arrays (640×512 diodes on $24 \mu\text{m}$ pitch) per layer or larger numbers of smaller arrays, and was comparable with the uniformity achieved by *Edwall* [15.76]. In the past, other workers have established MOVPE reactor designs capable of large-area uniform growth of CMT on 3 inch wafers [15.76, 77] but these activities have now ceased as attention has focused on MBE growth techniques.

The usual methods for determining depth uniformity of a CMT layer are sharpness of the infrared absorption edge and SIMS depth profiles (in particular looking at the Te^{125} secondary ion). The results from these techniques indicate that the IMP structure is fully diffused for IMP periods of the order of 1000 \AA and growth temperatures in the range of 350 to 400°C unless the surface becomes faceted during growth, when microinhomogeneities may occur [15.78].

15.3.1 Substrate Type and Orientation

The search for the correct substrate material and orientation has been a major area of research in CMT because it is a limiting factor in the quality of the epilayers. Essentially, there are two categories of substrates: (i) lattice-matched II–VI substrates and (ii) non-lattice-matched ‘foreign substrates’. Examples of the former are CdZnTe and CdSeTe, where the alloy compositions are tuned to the lattice parameter of the epilayer. Non-lattice-matched substrates include GaAs, Si and sapphire. The lattice mismatches can be up to 20% but, remarkably, heteroepitaxy is still obtained. The need for a ternary substrate to avoid substantial numbers of misfit dislocations has made the development of the CdTe-based substrate more complex. The small mismatch with CdTe substrates (0.2%) is sufficient to increase the dislocation density to greater than 10^6 cm^{-2} , comparable with some layers on CdTe-buffered GaAs, where the mismatch is 14% [15.76]. An additional problem encountered with the lattice-matched substrates is the lamella twins that form on (111) planes in Bridgman-grown crystals. It is possible to cut large ($4 \times 6 \text{ cm}$) (111)-oriented substrates parallel to the twins, but (100)

substrates have a much lower yield. MOVPE-grown CdTe also twins on the (111), and these twins will propagate through an entire structure. However, CMT growth on the (111)B face is very smooth for layers up to $20 \mu\text{m}$ thick, which is adequate for infrared detector structures.

The majority of MOVPE growth has concentrated on orientations close to (100), normally with a misorientation to reduce the size of macrodefects, known as hillocks or pyramids. Large Te precipitates can intersect the substrate surface and nucleate macrodefects. In a detailed analysis of the frequency and shapes of defects on different misorientations, it was concluded [15.79] that the optimum orientation was (100) $3-4^\circ$ towards the (111)B face. The presence of macrodefects is particularly critical for focal plane arrays, where they cause one or more defective pixels per defect. An alternative approach has been to use the (211)B orientation. In this case the surface appears to be free of macrodefects and is sufficiently misoriented from the (111) to avoid twinning. Dislocation densities of 10^5 cm^{-2} have been measured in CMT grown by IMP onto CdZnTe (211)B substrates [15.76], and diffusion-limited detectors have been fabricated using this orientation [15.80].

The alternative lattice-mismatched substrates were investigated as a more producible alternative to the variable quality of the CdTe family of substrates. As the CMT arrays must be cooled during operation, there is the risk that the differential thermal contraction between the substrate and multiplexer will break some of the indium contacting columns. The ideal substrate from this point of view is of course silicon, but the initial quality of heteroepitaxy with 20% lattice mismatch was poor. One of the most successful alternative substrate technologies has been the Rockwell PACE-I (producible alternative to CdTe epitaxy) which uses *c*-plane sapphire with a CdTe buffer layer grown by MOVPE and a CMT detector layer grown by LPE. The sapphire substrates absorb above $6 \mu\text{m}$ and can only be used for the $3-5 \mu\text{m}$ waveband. However, even with careful substrate preparation, a buffer layer thicker than $5 \mu\text{m}$ is needed to avoid contamination of the active layer.

GaAs has been the most extensively used alternative substrate, which has been successfully used to reduce the macrodefect density to below 10 cm^{-2} , and X-ray rocking curve widths below 100 arcs have been obtained [15.81]. Due to the large lattice mismatch, the layer nucleates with rafts of misfit dislocations that relieve any strain. The main cause of X-ray rocking curve broadening is the tilt associated with a mosaic structure that arises from the initial island growth.

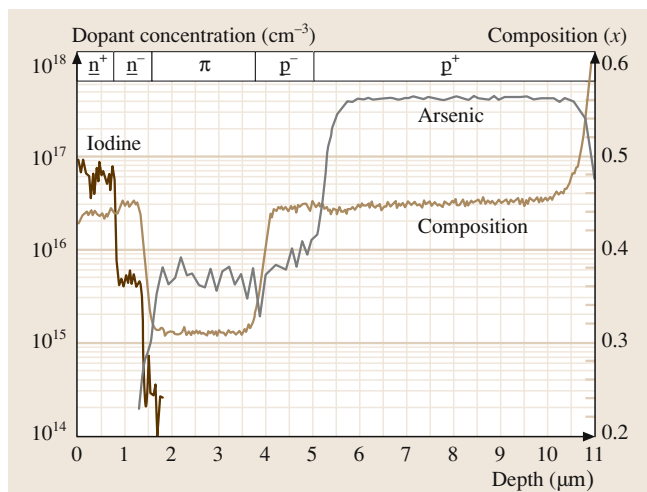


Fig. 15.8 SIMS profile of a $p^+p^-\pi n^-n^+$ device structure grown on a 3 inch wafer. (After [15.74])

Both CdTe and GaAs have the zinc blende structure but because the lattice mismatch is so large there is a better match with the orientation relationship (111)|| (100). The growth orientation of a CdTe epilayer can be either (100) or (111) according to the nucleation conditions. Tellurium-rich conditions appear to favor the (111) nucleation but temperature and choice of precursors also play an important role. A method for ensuring that the (100) orientation is achieved (thus avoiding twin formation) is to grow a thin nucleation layer of ZnTe that has a lattice parameter between that of CdTe and GaAs.

The majority of the MOVPE growth onto silicon substrates has used a GaAs buffer layer to step the change in lattice parameter between silicon and CdTe. By growing a 12 μm -thick CdTe buffer layer, the X-ray rocking curve width is reduced to 120 arcs, larger than the best values obtained on GaAs substrates [15.82]. X-ray topography reveals the same type of mosaic structure as for CdTe on bulk GaAs substrates. The advantage of thermal expansion match with the silicon multiplexer has been demonstrated with a midwave 256×256 array where the entire structure was grown by MOVPE [15.83]. Hybrid arrays of 480×640 elements have also been made, where MOVPE-grown CdZnTe/GaAs/Si was used as a substrate for LPE growth of the active CMT structure [15.65].

15.3.2 Doping

The earliest FPAs were made in as-grown, undoped MOVPE layers where the p-type nature was due to

metal vacancies in the layers. However, for current and future FPAs, the more complex devices such as p/n heterostructures require extrinsic doping and low-temperature annealing in a mercury vapor to remove metal vacancies. Activation of donors is relatively easy using either Group III metals or Group VII halogens. Indium has been used for doping concentrations from $< 10^{15} \text{ cm}^{-3}$ up to 10^{18} cm^{-3} . The low concentrations are limited by residual donors and acceptors, with the Hall mobility being a very sensitive measure of compensation. For the p/n structures, the n-type side of the junction is the absorber layer and therefore minority carrier lifetime and diffusion length are also important. An alternative n-type doping approach is to use a halogen such as iodine or chlorine that substitutes on the Group VI sites. The first use of iodine doping [15.84] used a crystalline iodine source, where the vapor was transported via hydrogen flow and a separate injector. The main advantage of iodine over indium is the slower diffusion, which enables more abrupt doped structures to be grown. A liquid source, ethyl iodide (EtI), was used by Mitra et al. [15.85] to dope over a range from 10^{15} to 10^{18} cm^{-3} , with no apparent memory. Current work on doped heterostructures [15.74] uses isobutyl iodide (IBI), as levels over a more useful range of (mid- 10^{14} to mid- 10^{17} cm^{-3}) can be achieved with this source donor.

Acceptor doping, as in LPE processes, proved to be more problematic in MOVPE than donor doping. The first attempts at arsenic doping used AsH_3 that would normally be very stable at growth temperatures of 350°C . It was proposed by Capper et al. [15.86] that the AsH_3 forms an adduct with DMCD in the vapor and causes the deposition of As at a low temperature. It was found that the doping incorporation was much higher in the CdTe layers and that activation of the dopant could be changed by adjusting the VI/II ratio during growth. A second-generation acceptor dopant source (tris-dimethyl amino arsine (DMAAs)) is now the preferred way to introduce As into layers. Multilayer, fully doped heterostructures have been grown [15.74, 87] which demonstrate higher operating temperature performance due to Auger suppression techniques. These heterostructures were grown to investigate the device performance in the 3–5 μm and 8–12 μm IR bands. Figure 15.8 shows a SIMS profile of a 3–5 μm device structure grown on a 3-inch-diameter GaAs substrate. The excellent control over the As concentration allows the active region to be doped at $4\text{--}5 \times 10^{15} \text{ cm}^{-3}$ in the so-called ‘ π -layer’ of the heterostructure design. The iodine profile shows well-defined transitions that

help define the junction depth. Such layers have been fabricated into mesa test and 2-D arrays. These have been indium-bumped onto either test array carriers or Si multiplexers. This material represents the current state-of-the-art in MOVPE.

15.3.3 In Situ Monitoring

The use of optical in situ monitoring probes for MOVPE is helping to both elucidate the kinetic mechanisms and provide monitors suitable for feedback control. Although in situ monitoring is not widely used in CMT growth, there have been a number of notable examples

where progress has been made as a result of its application. The choice of monitoring technique depends on the film property that is of interest. For example, laser reflectometry will measure the bulk of the film, measuring film thickness, growth rate and some indication of film composition. Reflection difference spectroscopy (RDS) is sensitive to surface composition and will monitor surface reconstruction and adsorbate composition. Spectroscopic ellipsometry is very sensitive to film composition and is now becoming useful as an in situ monitor with fast multichannel analysis. A number of in situ characterization techniques have been outlined in a review [15.73].

15.4 Molecular Beam Epitaxy (MBE)

An alternative to MOVPE for the production of advanced multilayer heterostructures is growth by molecular beam epitaxy (MBE). Figure 15.9 shows a schematic of a simple molecular beam epitaxial system for the growth of semiconductors. Significant progress was made in CMT MBE technology in the 1990s [15.88–91]. Current CMT MBE technology offers low-temperature growth under an ultrahigh vacuum environment, in situ n-type and p-type doping, and control of composition, doping and interfacial profiles. All of these are essential for the growth of double-layer and multilayer structures for advanced FPAs and other novel devices. A typical system consists of eight interconnected chambers, including two growth chambers, two preparation chambers with integrated ion etching capabilities and entry chambers, an ESCALAB chamber, two buffer chambers and a transport chamber. Each growth chamber is equipped with dual Hg reservoirs, eight effusion cells and dual load locks for effusion cells to enable recharging of the materials without breaking the vacuum. The system is modular in nature and designed to accommodate sample sizes up to 3 inch in diameter.

Normally, growth of CMT is carried out at 180–190 °C on (211) CdZnTe substrates using Hg, CdTe and Te sources. Typically, the CdZnTe substrates are first degreased, then etched in Br-methanol solution, and then rinsed with methanol. The substrates are blown with dry nitrogen and mounted on holders, and then thermally cleaned at 350–400 °C under vacuum to remove oxides prior to growth [15.92]. Although MBE is essentially a nonequilibrium growth process, the vapor pressure of the constituent species over a CMT layer is governed

by thermodynamics. It has been shown [15.93] that the vapor pressure of Hg over a CMT layer at 170 °C is $\approx 7 \times 10^{-4}$ mbar and 8 mbar for the Te-rich and Hg-rich sides of the phase boundary, respectively. The value of the Hg vapor pressure on the Te-rich side of the phase boundary is similar to the beam equivalent pressure of

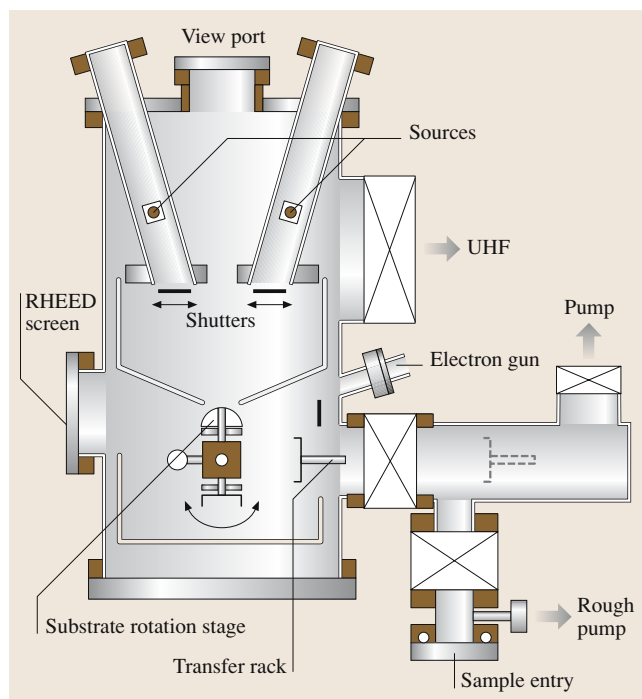


Fig. 15.9 Schematic of a simple molecular beam epitaxial system for the growth of semiconductors

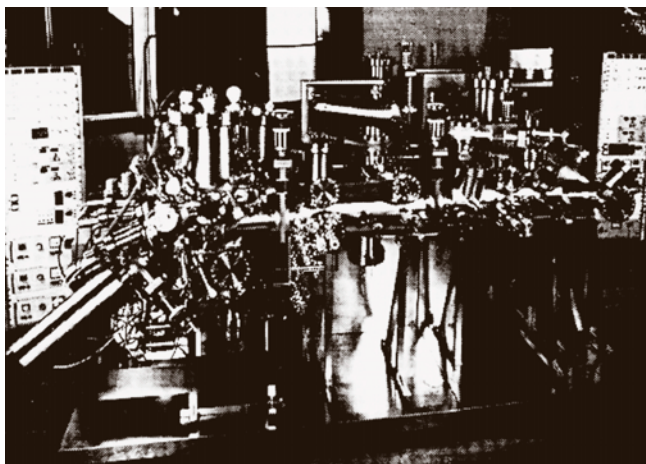


Fig. 15.10 CMT MBE growth facility. (After O.K. Wu et al., in [15.4] p. 97)

Hg employed for the growth of CMT by MBE, suggesting that CMT growth occurs on the Te-rich side of the phase boundary, and RHEED studies confirm this. Although Hg is the more mobile species, which is likely to attach at step edges, sufficient mobility of the Cd species is critical for the growth of high-quality films.

The CMT alloy composition can be readily varied by choosing the appropriate beam-flux ratio. Over the range from $x = 0.2 - 0.50$, excellent control of composition can be achieved readily by varying the CdTe source flux with a constant flux of Hg at 3×10^{-4} mbar and Te at 8×10^{-7} mbar during the MBE growth [15.91].

The most widely used n-type dopant for CMT alloys during MBE growth is indium [15.94, 95]. The In concentration can be varied from 2×10^{15} to $5 \times 10^{18} \text{ cm}^{-3}$ by adjusting the In cell temperature ($450 - 700^\circ\text{C}$) with no evidence of a memory effect. The doping efficiency of In was almost 100%, evident from the Hall measurement and secondary ion mass spectrometry (SIMS) data, for carrier concentrations $< 2 \times 10^{18} \text{ cm}^{-3}$. As in several bulk growth techniques, and in LPE and MOVPE processes, donor doping is seen to be much easier than acceptor doping in MBE growth.

A critical issue when growing advanced CMT structures is the ability to grow high-quality p-type materials in situ. As, Sb, N, Ag and Li have all been used as acceptors during MBE growth of CMT, with varying degrees of success [15.96]. Most data available is centered on the use of arsenic, and two approaches have been investigated. The first approach is based on photoassisted MBE to enable high levels of p-type As-doping of CdTe [15.95]. For As-doping during composition-

ally modulated structure growth, only the CdTe layers in a CdTe-CMT combination are doped, as in MOVPE growth. Since the CdTe does not contain Hg vacancies, and is grown under cation-rich conditions, the As is properly incorporated onto the Te site and its concentration is proportional to the As flux. The structure then interdiffuses after annealing at high temperature to remove residual Hg vacancies, resulting in p-type, homogeneous CMT. The main disadvantage of this approach is that it requires a high-temperature anneal that results in reduced junction and interface control. An alternative approach is to use cadmium arsenide and correct Hg/Te ratios to minimize Hg vacancies during CMT growth [15.94]. As a result, the As is directed to the Group VI sublattice to promote efficient p-type doping. The main growth parameters that determine the properties of As-doped p-type CMT are the growth temperature and Hg/Te flux ratio. A comparison of the net hole concentration and the SIMS measurement indicates that the electrical activity of the As acceptors exceeds 60%.

Lateral compositional and thickness uniformity, evaluated by nine-point FTIR measurements, were performed on a $2.5 \times 2.5 \text{ cm}^2$ sample, and the results showed that the average alloy composition and thickness were $x = 0.219 \pm 0.0006$ and $t = 8.68 \pm 0.064 \mu\text{m}$, respectively [15.91].

The surface morphology of CMT layers is important from a device fabrication point of view. Scanning electron microscopy (SEM) studies indicate that surface morphology of MBE-grown CMT alloys is very smooth for device fabrication, except for occasional small undulations ($< 1 \mu\text{m}$). The excellent crystal quality of CMT layers grown by MBE is illustrated by X-ray rocking curve data for a LWIR double-layer heterojunction structure. The In-doped n-type (about $8 \mu\text{m}$ thick) base layer peak has a width of < 25 arcs and is indistinguishable from the CdZnTe substrate. Because the As-doped p-type cap layer is much thinner (about $2 \mu\text{m}$) and has a different alloy composition, its peak is broader (45 arcs), but the X-ray FWHM width still indicates high quality [15.91].

Other material properties such as minority carrier lifetime and etch pit density of the material are important for device performance. The lifetime of the photoexcited carriers is among the most important, since it governs the diode leakage current and the quantum efficiency of a detector. In the case of In-doped n-type layers ($x = 0.2 - 0.3$), results show that the lifetime ranges from $0.5 - 3 \mu\text{s}$ depending upon the x value and carrier concentration. Measured

lifetimes at 77 K approach the Auger recombination limit [15.97].

Extended defects including dislocations, pinholes, particulates, inclusions, microtwins, precipitates and inhomogeneities occur over lengths of several micrometers and are often observable by selective etching and microscopy [15.59]. Optical microscopy studies of etched MBE-grown CMT base layers indicate that the EPD of device-quality alloys is on the order of $2 \times 10^5 \text{ cm}^{-2}$, which is in good agreement with the lifetime data and comparable to the best LPE materials for device applications.

15.4.1 Double-Layer Heterojunction Structures

Two approaches have been used to achieve these structures, namely ion implantation and in situ doping, in CMT MBE [15.96]. The advantage of the in situ doping approach is that it is a simple layer-by-layer growth process and so it is relatively easy to grow multilayer structures. Reproducible p-type doping with low defect density is difficult to achieve, and it requires very stringent passivation for mesa structures. However, high-performance CMT 64×64 and 128×128 FPAs for MWIR and LWIR infrared detection using CMT double-layer heterojunction structures (DLHJ) grown by MBE have been demonstrated [15.98]. The performances of detectors fabricated from MWIR and LWIR layers compare very well to those from the LPE-based production process.

15.4.2 Multilayer Heterojunction Structures

In recent years, as infrared technology has continued to advance, the demand for multispectral detectors has grown. The original growth process developed for the DLHJ structures was applied and extended for the growth of n-p-p-n multilayer heterojunction (MLHJ) structures for two-color detector applications [15.99]. A SIMS profile of an MLHJ structure confirms that the in situ doping process is able to incorporate $> 2 \times 10^{18} \text{ cm}^{-3}$ As and control the n-type doping at $1\text{--}2 \times 10^{15} \text{ cm}^{-3}$ for both absorbing layers [15.100]. X-ray rocking curve measurement indicates that the FWHM is about 50 arcs for this structure, suggesting high-quality material. This dual-band detector is basically a four-layer structure with two p-n junctions, one for each spectral band. The top and bottom layers are the IR absorbers while the middle two layers are transparent. The spectral re-

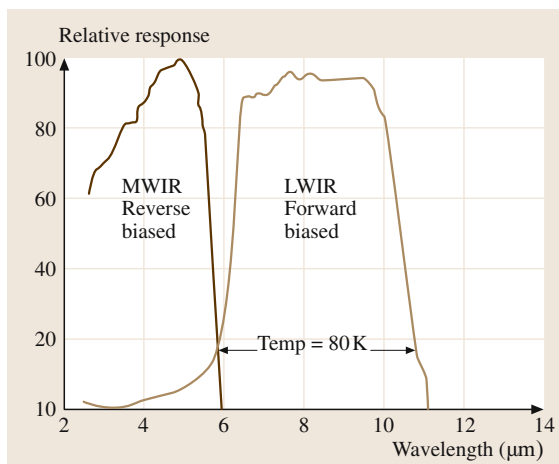


Fig. 15.11 The spectral response data for the MW-LWIR two-color detectors fabricated from a *n-p-p-n* multilayer heterojunction structure. The data indicate that the MWIR is detected in negative bias mode and LWIR is detected under forward bias conditions, as expected. (After O.K. Wu et al., in [15.4] p. 97)

sponse of this MW/LW two-color detector is shown in Fig. 15.11. This preliminary demonstration indicated that CMT MBE in situ doping technology is capable of growing multilayer structures for advanced IRFPAs.

15.4.3 CMT and CdZnTe Growth on Silicon

The move to using Si substrates for the epitaxial growth of CMT infrared IRFPAs is motivated by several important technological factors [15.91]. Primarily there is the need to establish a thermal expansion match between the Si readout electronics chip and the array of CMT infrared detectors. As the size of such CMT hybrid IRFPAs increases, long-term thermal cycle reliability can be compromised by the thermal expansion mismatch between Si and bulk CdZnTe substrates. One approach to solving this problem requires the fabrication of the CMT array on a Si substrate, rather than bulk CdZnTe, so that the array's thermal expansion is constrained to match that of the Si readout chip. Considerable research activity [15.101–104] has been directed towards achieving epitaxial deposition of high crystalline quality CMT films on Si substrates. Beyond their thermal expansion match to the readout electronics ship, Si substrates for IRFPAs offer additional advantages when compared with CdZnTe substrates; for example, they are more readily available in larger sizes with superior mechan-

ical strength and at substantially lower cost. The use of Si substrates also avoids potential problems associated with outdiffusion of fast-diffusing impurities, such as Cu, that has been identified [15.105] as a recurring problem with CdZnTe substrates. Finally, development of the technology for epitaxial growth of CMT on Si will ultimately be a requisite technology should monolithic integration of IR detector and readout electronics on a single Si chip become a goal of future IRFPA development.

However, the most serious technical challenge faced when fabricating device-quality epitaxial layers of CMT on Si is the reduction in the density of threading dislocations that results from the accommodation of the 19% lattice constant mismatch and the large difference in thermal expansion coefficients between Si and CMT. Dislocation density is known to have a direct effect on IR detector performance [15.58], particularly at low temperature. All efforts to fabricate CMT IR detectors on Si substrates have relied upon the prior growth of CdZnTe buffer layers on Si. Growth of $\approx 5\text{ }\mu\text{m}$ of CdZnTe is required to allow dislocation annihilation processes to decrease the dislocation density to low 10^6 cm^{-2} [15.104].

As an additional step, initiation layers of ZnTe have been used to facilitate parallel MBE deposition of CdTe(001) on Si(001) [15.98, 102]; ZnTe nucleation layers are also commonly used for the same purpose for the growth of CdZnTe on GaAs/Si substrates by other vapor-phase techniques [15.106]. CdTe(001) films with rocking curves as narrow as 78 arcsec and

EPD of $1\text{--}2\times 10^6\text{ cm}^{-2}$ have been demonstrated with this technique. Both (111)- and (001)-oriented MBE CdTe/Si substrates have been used as the basis for demonstrating LPE-grown CMT detectors [15.98, 101]. The (001)-oriented CdTe/Si films have been used in demonstrations of 256×256 CMT hybrid arrays on Si [15.98].

Current state-of-the-art MBE material on four-inch-diameter Si substrates has been discussed by Varesi et al. [15.107, 108]. Dry etching is used to produce array sizes of 128×128 and 1024×1024 with performances equivalent to LPE material. Similar material, grown on CdZnTe this time, by another group [15.109] in the MW and SW regions is used in astronomical applications (see [15.110]).

Other recent applications of MBE-grown CMT include very long wavelength arrays (onto (211)B CdZnTe substrates) by Philips et al. [15.111], two-color (MW $4.5\text{ }\mu\text{m}$ /SW $2.5\text{ }\mu\text{m}$) arrays of 128×128 diodes [15.69, 112], gas detectors in the $2\text{--}6\text{ }\mu\text{m}$ region [15.113] and $1.55\text{ }\mu\text{m}$ avalanche photodiodes using Si substrates [15.66]. All of these applications demonstrate the versatility of the CMT MBE growth technique.

One final point to note about the current devices being researched in MOVPE and MBE (and to a lesser extent LPE) processes is that growth is no longer of single layers from which the detector is made; instead the materials growers are actually producing the device structures within the grown layer. This is particularly true of the fully doped heterostructures grown by MOVPE and MBE shown in Figs. 15.8 and 15.11.

15.5 Alternatives to CMT

Rogalski [15.1, 114] has provided details about several Hg-based alternatives to CMT for infrared detection. He concludes that only HgZnTe and HgMnTe are serious candidates from the range of possibilities. Theoretical considerations of Sher et al. [15.115] showed that the Hg–Te bond is stabilized by the addition of ZnTe, unlike the destabilization that occurs when CdTe is added, as in CMT.

The pseudobinary phase diagram of HgZnTe shows even more separation of solidus and liquidus than the equivalent for CMT, see Fig. 15.12. This leads to large segregation effects, large composition variations for small temperature changes, and the high Hg vapor pressure presents the usual problems of containment. HgTe and MnTe are not completely miscible over the entire

range; the single-phase region is limited to $\lesssim x \lesssim 0.35$. The solidus–liquidus separation in the HgTe–MnTe pseudobinary is approximately half that in CMT, so for equivalent wavelength uniformity requirements any crystals of the former must be much more uniform than CMT crystals.

Three methods: Bridgman, SSR and THM are the most popular ones for the bulk growth of HgZnTe and HgMnTe. The best quality crystals have been produced by THM [15.116], with uniformities of ± 0.01 in both the axial and radial directions for HgZnTe. For HgMnTe, Bodnaruk et al. [15.117] produced crystals of $0.04 < x < 0.2$ with uniformities of ± 0.01 and ± 0.005 in the axial and radial directions, while Gille et al. [15.118] grew $x = 0.10$ crystals with ± 0.003 along

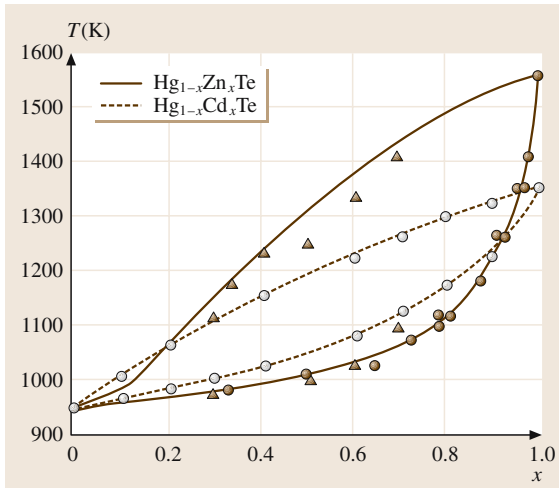


Fig. 15.12 HgTe-ZnTe and HgTe-CdTe pseudobinary phase diagrams. (After [15.120])

a 16 mm-diameter slice. *Becla et al.* [15.119] achieved decreased radial variations by applying a magnetic field.

These bulk crystals are suitable for some device architectures, but higher performance is obtained in more sophisticated structures that can only be made by epitaxial techniques. The advantages are the usual ones of lower growth temperature, lower Hg vapor pressure, shorter growth times and less precipitation with improved uniformity of composition. These features have been discussed in detail and descriptions of the various epitaxial processes given by *Rogalski* [15.114, 120, 121]. *LPE* growth of HgCdZnTe and HgCdMnTe by *Uchino* and *Takita* [15.122] showed that incorporation of Zn or Mn improved the uniformity of composition. HgMnTe films have been grown by *MOCVD* using the *IMP* technique, with both n- and p-type layers being produced. As

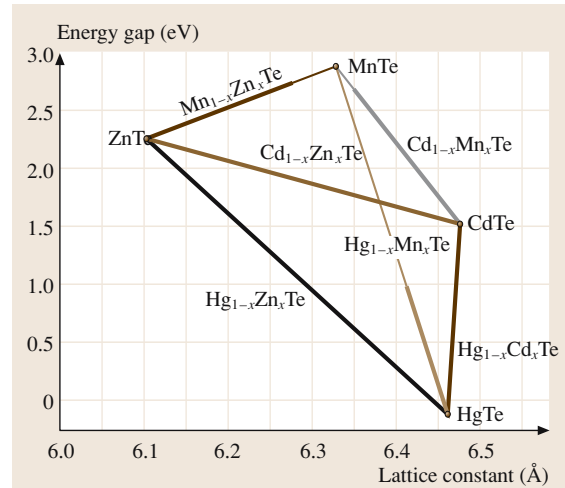


Fig. 15.13 Energy gap versus lattice parameter in Hg-based ternary alloy systems at room temperature. The bold lines indicate ranges of x for which homogeneous crystal phases form. (After A. Rogalski, in [15.6] p. 377)

in all epitaxial growth techniques, there is a need for suitable (nearly lattice-matched) large-area single-crystal substrates. Figure 15.13 shows the lattice constants of the Hg-based alloy systems at room temperature. There are larger differences between the lattice parameters of HgTe and ZnTe and HgTe and MnTe than between HgTe and CdTe. At first sight it would appear that CdMnTe would be a suitable substrate system for these ternary alloys, but CdMnTe crystals grown by the Bridgman process are highly twinned and therefore unsuitable for use as substrates.

The device aspects of these alternative ternary systems are discussed in the chapter by Baker in this Handbook.

References

- 15.1 A. Rogalski: *IR Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C. T. Elliott (Kluwer, Boston 2000)
- 15.2 M. Reine: *Encyclopedia of Modern Optics* (Academic, London 2002) p. 392
- 15.3 I. M. Baker: *Handbook of Infrared Detection Technologies*, ed. by M. Henini, M. Razeghi (Elsevier, Oxford 2003) Chap. 8
- 15.4 P. Capper (Ed.): *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications* (Chapman & Hall, London 1997)
- 15.5 P. Capper (Ed.): *Properties of Narrow Gap Cadmium-Based Compounds*, EMIS Datarev. Ser. (IEE, London 1994)
- 15.6 P. Capper, C. T. Elliott (Eds.): *IR Detectors and Emitters: Materials and Devices* (Kluwer, Boston 2000)
- 15.7 P. Capper (Ed.): *Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester 2005)
- 15.8 W. F. H. Micklethwaite: *Semicond. Semimet.* **18**, 48 (1981) Chap.3

- 15.9 P. W. Kruse: *Semicond. Semimet.* **18**, 1 (1981) Chap.1
- 15.10 R. Triboulet, T. Nguyen Duy, A. Durand: *J. Vac. Sci. Technol. A* **3**, 95 (1985)
- 15.11 W. E. Tennant, C. Cockrum, J. Gilpin, M. A. Kinch, M. B. Reine, R. P. Ruth: *J. Vac. Sci. Technol. B* **10**, 1359 (1992)
- 15.12 T. C. Harman: *J. Electron. Mater.* **1**, 230 (1972)
- 15.13 A. W. Vere, B. W. Straughan, D. J. Williams: *J. Cryst. Growth* **59**, 121 (1982)
- 15.14 L. Colombo, A. J. Syllaios, R. W. Perlaky, M. J. Brau: *J. Vac. Sci. Technol. A* **3**, 100 (1985)
- 15.15 R. K. Sharma, V. K. Singh, N. K. Mayyar, S. R. Gupta, B. B. Sharma: *J. Cryst. Growth* **131**, 565 (1987)
- 15.16 L. Colombo, R. Chang, C. Chang, B. Baird: *J. Vac. Sci. Technol. A* **6**, 2795 (1988)
- 15.17 J. Ziegler: US Patent 4,591,410 (1986)
- 15.18 W. M. Higgins, G. N. Pultz, R. G. Roy, R. A. Lancaster: *J. Vac. Sci. Technol. A* **7**, 271 (1989)
- 15.19 J. H. Tregilgas: *Prog. Cryst. Growth Charact.* **28**, 57 (1994)
- 15.20 P. Capper, J. Harris, D. Nicholson, D. Cole: *J. Cryst. Growth* **46**, 575 (1979)
- 15.21 P. Capper: *Prog. Cryst. Growth Charact.* **28**, 1 (1994)
- 15.22 A. Yeckel, and J.J. Derby: Paper given at 2002 US Workshop on Physics and Chemistry of II–VI Materials, San Diego, USA (2002)
- 15.23 P. Capper, and J.J.G. Gosney: U.K. Patent 8115911 (1981)
- 15.24 P. Capper, C. Maxey, C. Butler, M. Grist, J. Price: *Mater. Electron. Mater. Sci.* **15**, 721 (2004)
- 15.25 R. Triboulet: *Prog. Cryst. Growth Charact.* **28**, 85 (1994)
- 15.26 Y. Nguyen Duy, A. Durand, J. Lyot: *Mater. Res. Soc. Symp. Proc* **90**, 81 (1987)
- 15.27 A. Durand, J. L. Dessus, T. Nguyen Duy, J. Barbot: *Proc. SPIE* **659**, 131 (1986)
- 15.28 P. Gille, F. M. Kiessling, M. Burkert: *J. Cryst. Growth* **114**, 77 (1991)
- 15.29 P. Gille, M. Pesia, R. Bloedner, N. Puhlman: *J. Cryst. Growth* **130**, 188 (1993)
- 15.30 M. Royer, B. Jean, A. Durand, R. Triboulet: French Patent No. 8804370 (1/4/1988)
- 15.31 R. U. Bloedner, P. Gille: *J. Cryst. Growth* **130**, 181 (1993)
- 15.32 B. Chen, J. Shen, S. Din: *J. Electron Mater.* **13**, 47 (1984)
- 15.33 D. A. Nelson, W. M. Higgins, R. A. Lancaster: *Proc. SPIE* **225**, 48 (1980)
- 15.34 C.-H. Su, G. Perry, F. Szofran, S. L. Lehoczky: *J. Cryst. Growth* **91**, 20 (1988)
- 15.35 R. R. Galazka: *J. Cryst. Growth* **53**, 397 (1981)
- 15.36 B. Bartlett, P. Capper, J. Harris, M. Quelch: *J. Cryst. Growth* **47**, 341 (1979)
- 15.37 A. Durand, J. L. Dessus, T. Nguyen Duy: *Proc. SPIE* **587**, 68 (1985)
- 15.38 P. Capper, C. D. Maxey, C. L. Jones, J. E. Gower, E. S. O'Keefe, D. Shaw: *J. Electron Mater.* **28**, 637 (1999)
- 15.39 M. A. Kinch: *Mater. Res. Soc. Symp. Proc.* **90**, 15 (1987)
- 15.40 R. Pratt, J. Hewett, P. Capper, C. Jones, N. Judd: *J. Appl. Phys.* **60**, 2377 (1986)
- 15.41 R. Pratt, J. Hewett, P. Capper, C. L. Jones, M. J. T. Quelch: *J. Appl. Phys.* **54**, 5152 (1983)
- 15.42 F. Grainger, I. Gale, P. Capper, C. Maxey, P. Mackett, E. O'Keefe, J. Gosney: *Adv. Mater. Opt. Electron.* **5**, 71 (1995)
- 15.43 A. W. Vere: *Proc. SPIE* **659**, 10 (1986)
- 15.44 D. J. Williams, A. W. Vere: *J. Vac. Sci. Technol. A* **4**, 2184 (1986)
- 15.45 J. H. Tregilgas, J. D. Beck, B. E. Gnade: *J. Vac. Sci. Technol. A* **3**, 150 (1985)
- 15.46 C. Genzel, P. Gille, I. Hahnert, F. M. Kiessling, P. Rudolph: *J. Cryst. Growth* **101**, 232 (1990)
- 15.47 P. Capper, C. Maxey, C. Butler, M. Grist, J. Price: *J. Cryst. Growth* **275**, 259 (2005)
- 15.48 T. Tung: *J. Cryst. Growth* **86**, 161 (1988)
- 15.49 T.-C. Yu, R. F. Brebrick: *Properties of Narrow Gap Cadmium-Based Compounds*, EMIS Datarev. Ser., ed. by P. Capper (IEE, London 1994) p. 55
- 15.50 M. G. Astles: *Properties of Narrow Gap Cadmium-based Compounds*, EMIS Datarev. Ser., ed. by P. Capper (IEE, London 1994) p. 1
- 15.51 T. Tung, L. V. DeArmond, R. F. Herald: *Proc. SPIE* **1735**, 109 (1992)
- 15.52 P. W. Norton, P. LoVecchio, G. N. Pultz: *Proc. SPIE* **2228**, 73 (1994)
- 15.53 G. H. Westphal, L. Colombo, J. Anderson: *Proc. SPIE* **2228**, 342 (1994)
- 15.54 D. W. Shaw: *J. Cryst. Growth* **62**, 247 (1983)
- 15.55 L. Colombo, G. H. Westphal, P. K. Liao, M. C. Chen, H. F. Schaake: *Proc. SPIE* **1683**, 33 (1992)
- 15.56 I. B. Baker, G. J. Crimes, J. Parsons, E. O'Keefe: *Proc. SPIE* **2269**, 636 (1994)
- 15.57 P. Capper, E. S. O'Keefe, C. D. Maxey, D. Dutton, P. Mackett, C. Butler, I. Gale: *J. Cryst. Growth* **161**, 104 (1996)
- 15.58 R. S. List: *J. Electron. Mater.* **22**, 1017 (1993)
- 15.59 S. Johnson, D. Rhiger, J. Rosbeck: *J. Vac. Sci. Technol. B* **10**, 1499 (1992)
- 15.60 M. Yoshikawa: *J. Appl. Phys.* **63**, 1533 (1988)
- 15.61 P. Capper: *J. Vac. Sci. Technol. B* **9**, 1667 (1991)
- 15.62 C. A. Cockrum: *Proc. SPIE* **2685**, 2 (1996)
- 15.63 W. A. Radford, R. E. Kvaas, S. M. Johnson: *Proc. IRIS Specialty Group on Infrared Materials* (IRIS, Menlo Park 1986)
- 15.64 K. J. Riley, A. H. Lockwood: *Proc. SPIE* **217**, 206 (1980)
- 15.65 S. M. Johnson, J. A. Vigil, J. B. James: *J. Electron. Mater.* **22**, 835 (1993)
- 15.66 T. DeLyon, A. Hunter, J. Jensen, M. Jack, V. Randall, G. Chapman, S. Bailey, K. Kosai: Paper given at 2002 US Workshop on Physics and Chemistry of II–VI Materials, San Diego, USA (2002)
- 15.67 S. Johnson, J. James, W. Ahlgren: *Mater. Res. Soc. Symp. Proc.* **216**, 141 (1991)
- 15.68 P. R. Norton: *Proc. SPIE* **2274**, 82 (1994)

- 15.69 G. Destefanis, A. Astier, J. Baylet, P. Castelein, J. P. Chamonal, E. De Borniol, O. Gravand, F. Marion, J. L. Martin, A. Million, P. Rambaud, F. Rothan, J. P. Zanatta: *J. Electron. Mater.* **32**, 592 (2003)
- 15.70 T. F. Kuech, J. O. McCaldin: *J. Electrochem. Soc.* **128**, 1142 (1981)
- 15.71 S. J. C. Irvine, J. B. Mullin: *J. Cryst. Growth* **55**, 107 (1981)
- 15.72 A. C. Jones: *J. Cryst. Growth* **129**, 728 (1993)
- 15.73 S. J. C. Irvine, J. Bajaj: *Semicond. Sci. Technol.* **8**, 860 (1993)
- 15.74 C. D. Maxey, J. Camplin, I. T. Guilfooy, J. Gardner, R. A. Lockett, C. L. Jones, P. Capper: *J. Electron. Mater.* **32**, 656 (2003)
- 15.75 J. Tunnicliffe, S. Irvine, O. Dosser, J. Mullin: *J. Cryst. Growth* **68**, 245 (1984)
- 15.76 D. D. Edwall: *J. Electron. Mater.* **22**, 847 (1993)
- 15.77 S. Murakami: *J. Vac. Sci. Technol. B* **10**, 1380 (1992)
- 15.78 S. J. C. Irvine, D. Edwall, L. Bubulac, R. V. Gil, E. R. Gertner: *J. Vac. Sci. Technol. B* **10**, 1392 (1992)
- 15.79 D. W. Snyder, S. Mahajan, M. Brazil: *Appl. Phys. Lett.* **58**, 848 (1991)
- 15.80 P. Mitra, Y. L. Tyan, F. C. Case: *J. Electron. Mater.* **25**, 1328 (1996)
- 15.81 A. M. Kier, A. Graham, S. J. Barnett: *J. Cryst. Growth* **101**, 572 (1990)
- 15.82 S. J. C. Irvine, J. Bajaj, R. V. Gil, H. Glass: *J. Electron. Mater.* **24**, 457 (1995)
- 15.83 S. J. C. Irvine, E. Gertner, L. Bubulac, R. V. Gil, D. D. Edwall: *Semicond. Sci. Technol.* **6**, C15 (1991)
- 15.84 C. D. Maxey, P. Whiffin, B. C. Easton: *Semicond. Sci. Technol.* **6**, C26 (1991)
- 15.85 P. Mitra, Y. L. Tyan, T. R. Schimert, F. C. Case: *Appl. Phys. Lett.* **65**, 195 (1994)
- 15.86 P. Capper, C. Maxey, P. Whiffin, B. Easton: *J. Cryst. Growth* **97**, 833 (1989)
- 15.87 C. D. Maxey, C. J. Jones: *Proc. SPIE* **3122**, 453 (1996)
- 15.88 R. D. Rajavel, D. Jamba, O. K. Wu, J. A. Roth, P. D. Brewer, J. E. Jensen, C. A. Cockrum, G. M. Venzor, S. M. Johnson: *J. Electron. Mater.* **25**, 1411 (1996)
- 15.89 J. Bajaj, J. M. Arias, M. Zandian, D. D. Edwall, J. G. Pasko, L. O. Bubulac, L. J. Kozlowski: *J. Electron. Mater.* **25**, 1394 (1996)
- 15.90 J. P. Faurie, L. A. Almeida: *Proc. SPIE* **2685**, 28 (1996)
- 15.91 O. K. Wu, T. J. deLyon, R. D. Rajavel, J. E. Jensen: *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications*, ed. by P. Capper (Chapman & Hall, London 1997) p. 97
- 15.92 O. K. Wu, D. R. Rhiger: *Characterization in Compound Semiconductor Processing*, ed. by Y. Strausser, G. E. McGuire (Butterworth-Heinemann, London 1995) p. 83
- 15.93 T. Tung, L. Golonka, R. F. Brebrick: *J. Electrochem. Soc.* **128**, 451 (1981)
- 15.94 O. Wu, D. Jamba, G. Kamath: *J. Cryst. Growth* **127**, 365 (1993)
- 15.95 J. Arias, S. Shin, D. Copper: *J. Vac. Sci. Technol. A* **8**, 1025 (1990)
- 15.96 O. K. Wu: *Mater. Res. Soc. Symp. Proc.* **340**, 565 (1994)
- 15.97 V. Lopes, A. J. Syllaos, M. C. Chen: *Semicond. Sci. Technol.* **8**, 824 (1993)
- 15.98 S. M. Johnson, T. J. de Lyon, C. Cockrum: *J. Electron. Mater.* **24**, 467 (1995)
- 15.99 G. Kamath, and O. Wu: US Patent Number 5,028,561, July 1, 1991
- 15.100 O. K. Wu, R. D. Rajavel, T. J. deLyon: *Proc. SPIE* **2685**, 16 (1996)
- 15.101 F. T. Smith, P. W. Norton, P. Lo Vecchio: *J. Electron. Mater.* **24**, 1287 (1995)
- 15.102 J. M. Arias, M. Zandian, S. H. Shin: *J. Vac. Sci. Technol. B* **9**, 1646 (1991)
- 15.103 R. Sporken, Y. Chen, S. Sivananthan: *J. Vac. Sci. Technol. B* **10**, 1405 (1992)
- 15.104 T. J. DeLyon, D. Rajavel, O. K. Wu: *Proc. SPIE* **2554**, 25 (1995)
- 15.105 J. P. Tower, S. P. Tobin, M. Kestigian: *J. Electron. Mater.* **24**, 497 (1995)
- 15.106 N. Karam, R. Sudharsanan: *J. Electron. Mater.* **24**, 483 (1995)
- 15.107 J. B. Varesi, A. A. Buell, R. E. Bornfreund, W. A. Radford, J. M. Peterson, K. D. Maranowski, S. M. Johnson, D. F. King: *J. Electron. Mater.* **31**, 815 (2002)
- 15.108 J. B. Varesi, A. A. Buell, J. M. Peterson, R. E. Bornfreund, M. F. Vilela, W. A. Radford, S. M. Johnson: *J. Electron. Mater.* **32**, 661 (2003)
- 15.109 M. Zandian, J. D. Garnett, R. E. DeWames, M. Carmody, J. G. Pasko, M. Farris, C. A. Cabelli, D. E. Cooper, G. Hildebrandt, J. Chow, J. M. Arias, K. Vural, D. N. B. Hall: *J. Electron. Mater.* **32**, 803 (2003)
- 15.110 I. S. McLean: Paper given at 2002 US Workshop on Physics and Chemistry of II–VI Materials, San Diego, USA (2002)
- 15.111 J. D. Philips, D. D. Edwall, D. L. Lee: *J. Electron. Mater.* **31**, 664 (2002)
- 15.112 L. A. Almeida, M. Thomas, W. Larsen, K. Spariosu, D. D. Edwall, J. D. Benson, W. Mason, A. J. Stolz, J. H. Dinan: *J. Electron. Mater.* **31**, 669 (2002)
- 15.113 J. P. Zanatta, F. Noel, P. Ballet, N. Hdadach, A. Million, G. Destefanis, E. Mottin, E. Picard, E. Hadji: *J. Electron. Mater.* **32**, 602 (2003)
- 15.114 A. Rogalski: *New Ternary Alloy Systems for Infrared Detectors* (SPIE Optical Engineering, Bellingham 1994)
- 15.115 A. Sher, A. B. Chen, W. E. Spicer, C. K. Shih: *J. Vac. Sci. Technol. A* **3**, 105 (1985)
- 15.116 R. Triboulet: *J. Cryst. Growth* **86**, 79 (1988)
- 15.117 O. A. Bodnaruk, I. N. Gorbatiuk, V. I. Kalenik: *Neorg. Mater.* **28**, 335 (1992)
- 15.118 P. Gille, U. Rössner, N. Puhlmann: *Semicond. Sci. Technol.* **10**, 353 (1995)
- 15.119 P. Becla, J.-C. Han, S. Matakaf: *J. Cryst. Growth* **121**, 394 (1992)

- [15.120](#) A. Rogalski: Prog. Quantum Electron. **13**, 299 (1989) [15.122](#) T. Uchino, K. Takita: J. Vac. Sci. Technol. A **14**, 2871 (1996)
- [15.121](#) A. Rogalski: Infrared Phys. **31**, 117 (1991)

16. Wide-Bandgap II–VI Semiconductors: Growth and Properties

Wide-bandgap II–VI compounds have been applied to optoelectronic devices, especially light-emitting devices in the short-wavelength region of visible light, because of their direct gap and suitable bandgap energies. Many methods have been extensively applied to grow high-quality films and bulk single crystals from the vapor and liquid phases.

This chapter firstly discusses the basic properties and phase diagrams of wide-bandgap II–VI compounds such as ZnS, ZnO, ZnSe, ZnTe, CdSe and CdTe. Then the growth methods and recent progress in films and bulk crystal growth are reviewed. In the epitaxial growth methods, the focus is on liquid-phase epitaxy (LPE), vapor-phase epitaxy (VPE) containing conventional VPE, hot-wall epitaxy (HWE), metalorganic chemical vapor deposition (MOCVD) or metalorganic phase epitaxy (MOVPE), molecular-beam epitaxy (MBE) and atomic-layer epitaxy (ALE). In bulk crystal growth, two typical growth methods,

16.1	Crystal Properties	326
16.1.1	Basic Properties	326
16.1.2	Phase Diagram.....	326
16.2	Epitaxial Growth	328
16.2.1	The LPE Technique	329
16.2.2	Vapor-Phase Epitaxy Techniques	330
16.3	Bulk Crystal Growth	333
16.3.1	The CVT and PVT Techniques	333
16.3.2	Hydrothermal Growth	336
16.3.3	Bridgman and Gradient Freezing (GF) Method	337
16.3.4	The Traveling Heater Method (THM)	338
16.3.5	Other Methods	338
16.4	Conclusions	339
	References	340

chemical/physical vapor transport (CVT/PVT) and Bridgman techniques, are introduced.

Wide-bandgap II–VI compounds are expected to be one of the most vital materials for high-performance optoelectronics devices such as light-emitting diodes (LEDs) and laser diodes (LDs) operating in the blue or ultraviolet spectral range. Additionally, the high ionicity of these compounds makes them good candidates for high electro-optical and electromechanical coupling. The basic promises of wide-bandgap materials can be found in Fig. 16.1.

Thin films were commonly grown using the conventional vapor-phase epitaxy (VPE) method for 60 s. With the development of science and technology, new and higher requirements arose for material preparation. For this reason, novel epitaxial growth techniques were developed, including hot-wall epitaxy (HWE) [16.1], metalorganic chemical vapor deposition (MOCVD) [16.2], molecular-beam epitaxy (MBE) [16.3], metalorganic molecular-beam epitaxy (MOMBE) [16.4] and atomic-layer epitaxy (ALE) [16.5]. Using these growth methods, film thickness can be controlled, and quality can be improved.

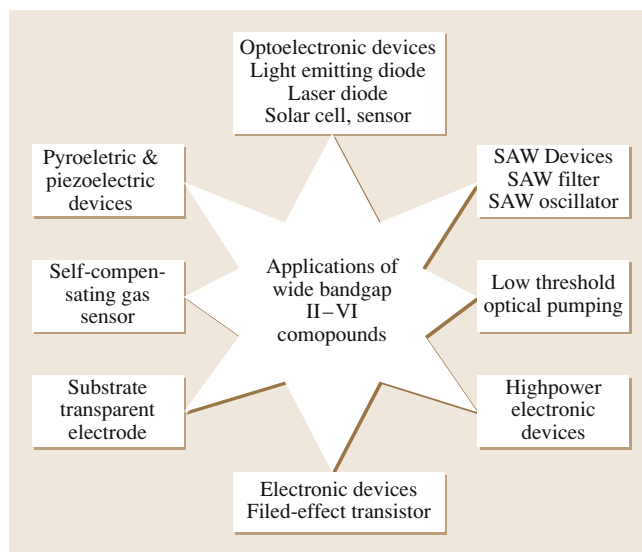


Fig. 16.1 Application of II–VI wide-bandgap compounds

On the other hand, basic research work into growing bulk crystals of wide-bandgap II–VI compounds has been carried out. Focus was put on high-purity, high-quality, large single crystals [16.6–10]. Since the electrical and optical properties of semiconductor compounds are drastically affected by impurities and native defects, purity and quality are very important for fundamental research and engineering application where they are used as substrates. Bulk single crystals of these wide-bandgap II–VI compounds have been grown from the vapor, liquid and solid phases. Vapor-phase growth includes chemical vapor transport (CVT) and physical vapor transport (PVT) methods; liquid-phase methods includes growth from the melt or solvent. Among these growth methods, melt growth is most suitable to produce sizable bulk crystals for relatively short growth duration. Growth methods for films and bulk crystals of wide-bandgap II–VI compounds are summarized in Fig. 16.2 and the details can be found in Sects. 16.2 and 16.3.

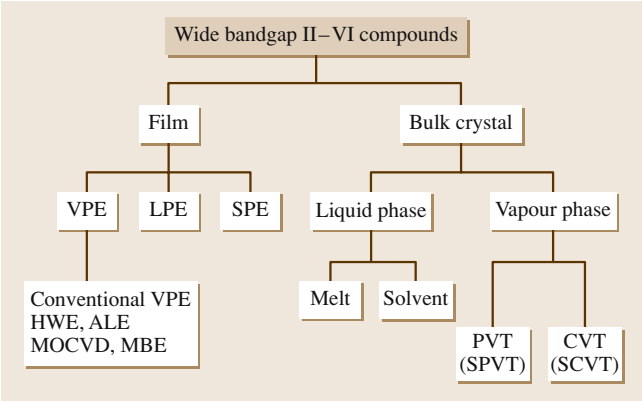


Fig. 16.2 Film and bulk-crystal growth techniques for II–VI wide-bandgap compounds

16.1 Crystal Properties

16.1.1 Basic Properties

Wide-bandgap compound semiconductors have higher melting points. Due to their high ionicity, the overheating phenomenon occurs when they are heated to their melting point. Owing to the higher vapor pressures at their melting points, it is difficult to grow bulk crystals from melt. On the other hand, it is easy to grow bulk crystals as well as their films from the vapor phase. Therefore,

This chapter firstly describes the physical and chemical properties of these wide-bandgap II–VI compounds, then reviews the growth techniques and introduces the main results in preparing film and bulk single crystals of ZnS, ZnO, ZnSe, ZnTe, CdTe, and so on.

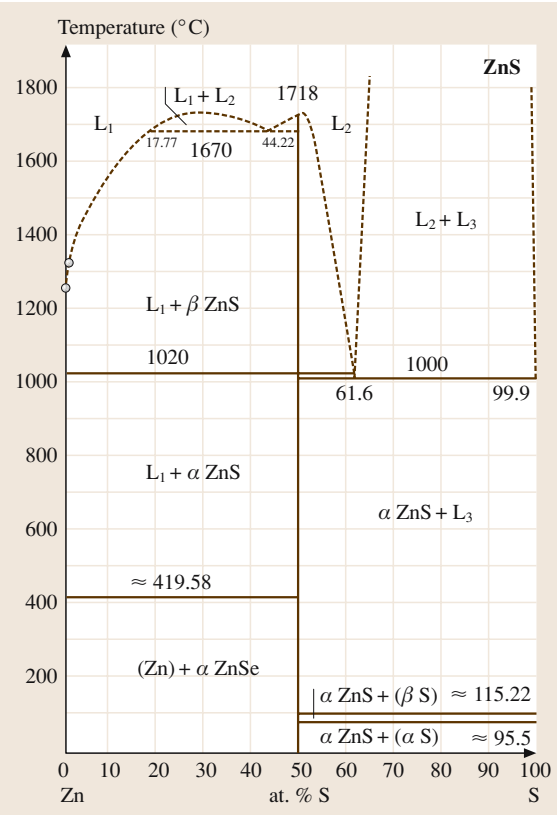


Fig. 16.3a–g Phase diagrams of some main wide-bandgap II–VI compounds. (a) ZnS, (b) ZnSe, (c) ZnSe phase diagram near the congruent point; (d) ZnTe; (e) CdSe; (f) CdTe; (g) CdTe phase diagram near the congruent point

before introducing film and bulk crystal growth, it is necessary to review their physical and chemical properties. Table 16.1 shows the properties of some of the main II–VI compound semiconductor materials [16.11–24].

16.1.2 Phase Diagram

It is necessary to understand the phase diagram to grow high-quality film and bulk single crystals. Figure 16.3

Table 16.1 Properties of some wide-bandgap II–VI compound semiconductors

Material Property	ZnS	ZnO	ZnSe	ZnTe	CdS	CdSe	CdTe
Melting point (K)	2038 (WZ, 150 atm)	2248	1797	1513	2023 (WZ, 100 atm)	1623	1370 (ZB)
Energy gap E_g at 300 K (eV)(ZB*/WZ*)	3.68/3.911	–/3.4	2.71/–	2.394	2.50/2.50	–/1.751	1.475
dE_g/dT ($\times 10^{-4}$ eV/K) ZB/WZ	4.6/8.5	–/9.5	4.0/–	5.5/–	–/5.2	–/4.6	5.4/–
Structure	ZB/WZ	WZ	ZB/WZ	ZB	WZ	WZ	ZB
Bond length (μm)	2.342 (WZ)	1.977 (WZ)	2.454 (ZB)	2.636 (ZB)	2.530 (ZB)	2.630 (ZB)	2.806 (ZB)
Lattice constant (ZB) a_0 at 300 K (nm)	0.541	–	0.567	0.610	0.582	0.608	0.648
ZB nearest-neighbor dist. at 300 K (nm)	0.234	–	0.246	0.264	0.252	0.263	0.281
ZB density at 300 K (g/cm^3)	4.11	–	5.26	5.65	4.87	5.655	5.86
Lattice constant (WZ) at 300 K (nm)							
$a_0 = b_0$	0.3811	0.32495	0.398	0.427	0.4135	0.430	–
c_0	0.6234	0.52069	0.653	0.699	0.6749	0.702	–
c_0/a_0	1.636	1.602	1.641	1.637	1.632	1.633	–
WZ density at 300 K (g/cm^3)	3.98	5.606	–	–	4.82	5.81	–
Symmetry ZB/WZ	C6me/F43m	–/C6me	–/F43m	–/F43m	C6me/F43m	C6me/F43m	–/–
Electron affinity χ (eV)			4.09	3.53	4.79	4.95	4.28
Stable phase(s) at 300 K	ZB & WZ	WZ	ZB	ZB	ZB & WZ	ZB & WZ	ZB
Solid–solid phase transition temperature (K)	1293	–	1698	–	–	403	1273(?)
Heat of crystallization ΔH_{LS} (kJ/mol)	44	62	52	56	58	45	57
Heat capacity C_p (cal/mol K)	11.0	9.6	12.4	11.9	13.2	11.8	–
Ionicity (%)	62	62	63	61	69	70	72
Equilibrium pressure at c.m.p. (atm)	3.7	–	1.0	1.9	3.8	1.0	0.7
Minimum pressure at m.p. (atm)	2.8	7.82	0.53	0.64	2.2	0.4–0.5	0.23
Specific heat capacity (J/gK)	0.469	–	0.339	0.16	0.47	0.49	0.21
Thermal conductivity ($\text{W cm}^{-1}\text{K}^{-1}$)	0.27	0.6	0.19	0.18	0.2	0.09	0.01
Thermo-optical coefficient (dn/dT)($\lambda = 10.6 \mu\text{m}$)	4.7	–	6.1	–	–	–	11.0
Electrooptical coefficient r_{41} (m/V) ($\lambda = 10.6 \mu\text{m}$)	2×10^{-12}	–	2.2×10^{-12}	4.0×10^{-12} ($r_{41} = r_{52}$ $= r_{63}$)	–	–	6.8×10^{-12}

m.p. – melting point; c.m.p. – congruent melting point; ZB – zinc blende; WZ – wurtzite

Table 16.1 (continued)

Material Property	ZnS	ZnO	ZnSe	ZnTe	CdS	CdSe	CdTe
Linear expansion coefficient (10 ⁻⁶ K ⁻¹) ZB/WZ	-/6.9	2.9/7.2	7.6/-	8.0/-	3.0/4.5	3.0/7.3	5.1/-
Poisson ratio	0.27		0.28				0.41
Dielectric constant $\epsilon_0/\epsilon_\infty$	8.6/5.2	8.65/4.0	9.2/5.8	9.3/6.9	8.6/5.3	9.5/6.2	2.27/-
Refractive index ZB/WZ	2.368/2.378	-/2.029	2.5/-	2.72/-	-/2.529	2.5/-	2.72/-
Absorption coeff. (including two surfaces) ($\lambda = 10.6 \mu\text{m}$)(cm ⁻¹)	≤ 0.15	-	1-2 × 10 ⁻³	-	≤ 0.007	≤ 0.0015	≤ 0.003
Electron effective mass (m^*/m_0)	-0.40	-0.27	0.21	0.2	0.21	0.13	0.11
Hole effective mass m_{dos}^*/m_0	-	-	0.6	circa 0.2	0.8	0.45	0.35
Electron Hall mobility (300) K for $n =$ lowish (cm ² /Vs)	165	125	500	340	340	650	1050
Hole Hall mobility at 300 K for $p =$ lowish (cm ² /Vs)	5	-	30	100	340	-	100
Exciton binding energy (meV)	36	60	21	10	30.5	15	12
Average phonon energy (meV) ZB/WZ	16.1/17.1	-	15.1/-	10.8/-	-/13.9	18.9/25.4	5.8/-
Elastic constant (10 ¹⁰ N/m ²)							
C ₁₁	1.01±0.05	-	8.10±0.52	0.72±0.01	-	-	5.57
C ₁₂	0.64±0.05	-	4.88±0.49	0.48±0.002	-	-	3.84
C ₄₄	0.42±0.04	-	4.41±0.13	0.31±0.002	-	-	2.095
Knoop hardness (N/cm ²)	0.18	0.5	0.15	0.13	-	-	0.10
Young's modulus	10.8 Mpsi	-	10.2 Mpsi	-	45 GPa	5 × 10 ¹¹ dyne/cm ²	3.7 × 10 ¹¹ dyne/cm ²

m.p. – melting point; c.m.p. – congruent melting point; ZB – zinc blende; WZ – wurtzite

shows the phase diagrams reported for ZnS [16.25], ZnSe [16.25, 26], ZnTe [16.25], CdSe [16.27] and CdTe [16.28, 29]. Although much work has been done, there some exact thermodynamic data are still

lacking, especially details close to the congruent point. Unfortunately, the phase diagram of ZnO is not available in spite of its growing importance in applications.

16.2 Epitaxial Growth

Epitaxial growth of wide-bandgap II–VI compounds was mainly carried out using liquid-phase epitaxy (LPE), or VPE. VPE includes several techniques, such as conventional VPE, hot-wall epitaxy (HWE), met-

alorganic chemical vapor deposition (MOCVD) or metalorganic phase epitaxy (MOVPE), molecular-beam epitaxy (MBE), metalorganic molecular-beam epitaxy (MOMBE) and atomic-layer epitaxy (ALE), etc. Each

Table 16.2 Strengths and weaknesses of several epitaxial growth techniques

LPE	Thermodynamic equilibrium growth Easy-to-use materials Low-temperature growth High purity Multiple layers Thickness control not very precise Poor surface/interface morphology	MOCVD	Gaseous reaction for deposition Precise composition Patterned/localized growth Potentially easier large-area multiple-wafer scale-up Low-temperature growth High-vapor-pressure materials growth allowed About 1 ML/s deposition rate Expensive equipment Safety precautions needed
HWE	Easy-to-use materials Low cost Thermodynamic equilibrium Hard to grow thick layers Thickness control not very precise	MBE and MOMBE	Physical vapor deposition Ultra-high-vacuum environment About 1 ML/s deposition rate In situ growth-front monitoring Precise composition Low growth rate Sophisticated equipment Limit for high-vapor-pressure materials growth (MBE)
VPE	Easy to operate Economic Thinner layers High growth rates Easier composition control High temperature (800–1000 °C)		
ALE	Gaseous reaction for deposition Low-temperature growth Precise composition Low growth rate Safety precautions needed		

of these methods has its advantages and disadvantages. They are summarized in Table 16.2.

In the case of hetero-epitaxy, the mismatch between substrate material and epitaxial layer affects the growing structure and quality of the epitaxial layer. The mismatch should be made as small as possible when choosing the pair of materials (substrate and epitaxial material). Furthermore, the difference between the thermal expansion coefficients of the pair of materials has to be considered to obtain high-quality epitaxial layer [16.30].

16.2.1 The LPE Technique

LPE growth occurs at near-thermodynamic-equilibrium conditions. There are two growth methods. The first is called equilibrium cooling, in which the saturated solution is in contact with the substrate and the temperature is lowered slowly, the solution becomes supersaturated; meanwhile a slow epitaxial growth on the substrate is ini-

tiated. The second is the step-cooling process, in which the saturated solution is cooled down a few degrees (5–20 K) to obtain a supersaturated solution. The substrate is inserted into the solution, which is kept at this cooled temperature. Growth occurs first due to the supersaturation, and will slow down and stop finally. For both techniques, if the substrate is dipped in sequence into several different melt sources, multiple layer structures can be grown. **LPE** can successfully and inexpensively grow homo- and heterostructures. As the growth is carried out under thermal equilibrium, an epilayer with a very low native defect density can be obtained.

The **LPE** method can be used to grow high-quality epilayers, such as ZnS [16.31], ZnSe [16.31, 32], ZnSSe [16.33], ZnTe [16.34], etc. *Werkhoven et al.* [16.32] grew ZnSe epilayers by **LPE** on ZnSe substrates in a low-contamination-level environment. In their study, the width of bound exciton lines in low-temperature photoluminescence spectra was used

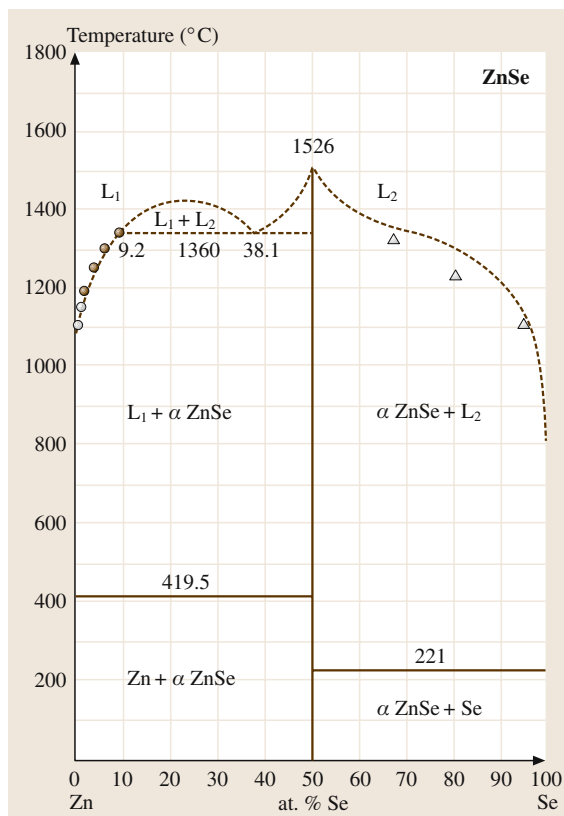


Fig. 16.4 A typical VPE growth system

to define the quality of the material, and the energy of the lines was used to identify trace impurities. The photoluminescence (PL) results showed that the ZnSe epitaxial layer has the high quality. The sharpest spectra occurred in layers grown rapidly on a previously grown buffer layer, indicating the importance of impurity out-diffusion from the substrate into the growing layer. The sharpness of these bound exciton lines indicates that the total concentration of electrically active impurities ($N_A + N_D$) was below 10^{17} cm^{-3} .

16.2.2 Vapor-Phase Epitaxy Techniques

Conventional VPE

A typical VPE growth system is shown in Fig. 16.4. In VPE growth, thin films are formed by the deposition of atoms from the vapor phase. There are two types of transport mechanisms for the source materials, physical vapor deposition (PVD) without any chemical reaction, and chemical vapor deposition (CVD), where the formation of the deposited film is the result of a chemical reaction

of the precursors on the substrate. In VPE growth, there are several important parameters, such as the source temperature, the substrate temperature, the flow rate of the carrier gas, the growth pressure, and so on. These determine the growth rate, composition, and crystallinity of the epitaxial layers.

The VPE technique is the most popular in semiconductor epitaxial growth. Since the vapor pressures of all wide-bandgap II–VI materials are high, their epitaxial layers can be grown by this method. As an example, high-quality ZnS single-crystal films have been grown on a Si substrate using hydrogen as a carrier gas [16.35, 36]. Furthermore, the Iida group [16.37] doped N and P into a ZnS epilayer and studied their behavior in details. The N and P were expected to compensate the native donor state and to result in an insulating material. The results showed that the doped acceptors N and P reduced the donor density and an insulating material was obtained. Later, this group was successful in preparing a p-type ZnS epilayer using NH_3 as an acceptor dopant [16.38]. Many efforts have been made to grow ZnSe epilayers on different substrates, especially on GaAs in the past two decades [16.39, 40]. p-type ZnSe was also obtained by this technique [16.41]. In addition, other compounds, ZnTe [16.42], CdS [16.43], CdSe [16.44], and CdTe [16.45], were also studied using this technique.

The HWE Technique

Hot-wall epitaxy [16.1] has proved to be a very successful growth method for II–VI compound epitaxial layers. Its principal characteristic is the growth of thin films under conditions near thermodynamic equilibrium. Compared with other VPE methods, the HWE technique has the advantages of low cost, simplicity, convenience, and relatively high growth rate. In particular, it can control deviation from stoichiometry during growth of an epilayer.

A schematic diagram of the improved HWE system is showed in Fig. 16.5; it consists of four independent furnaces. The source material placed at the middle is transported to the substrate. The region of the growth reactor between the source and substrate, called the hot wall, guarantees a nearly uniform and isotropic flux of molecules onto the substrate surface. To control the deviation from stoichiometry, the reservoir part is placed at the bottom with a constituent element.

HWE has been applied to growing II–VI compound epilayers, such as CdTe [16.46], CdS [16.47], CdSe [16.48], ZnTe [16.49], and also to producing heterostructures for laser and photovoltaic detector fab-

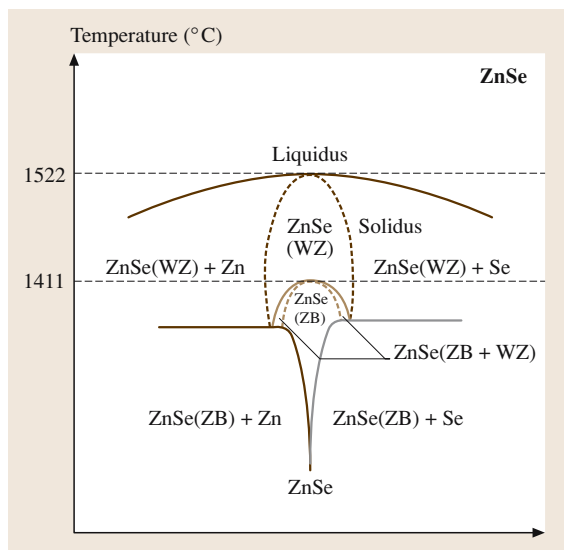


Fig. 16.5 Diagram of a typical HWE system growth chamber

rication [16.50]. Most research using HWE technique has focused on CdTe growth. Wang et al. [16.46] optimized growth conditions and grew high-quality CdTe epitaxial films using the HWE apparatus shown in Fig. 16.5. All the CdTe epilayers show mirror-like surfaces. Results from PL and X-ray diffraction (XRD) show that CdTe epilayers on GaAs suffer from a biaxial compressive stress, that this stress is rapidly relaxed within a thickness of about 5 μm , and that it remains in the epilayer up to a film thickness of 15 μm . Although this heterosystem has a 14.6% lattice mismatch and -26% thermal expansion mismatch at 300 K, high-quality CdTe epilayers, with a full-width half-maximum (FWHM) of 0.26 meV for bound-exciton emission lines in 4.2-K PL, about 90 arcs for (400) diffraction in four-crystal XRD spectrum, were prepared by selecting suitable growth conditions and epilayer thickness.

Recently, significant results have been achieved for CdTe/Si (111) epilayer growth by HWE. Lalev et al. [16.51] reported that high-quality CdTe (111) epilayers with A polarity were directly grown on hydrogen-terminated Si (111) without any preheating treatment. Through the originally designed two-step growth regime, the crystal quality of CdTe film was significantly improved, and the best FWHM value of 118 arcs from four-crystal rocking curves was obtained for a 5- μm -thick epilayer.

The MOCVD Technique

MOCVD or MOVPE is an improvement over conventional VPE. Since its introduction in 1968 [16.52], this technique has been established as one of the techniques for epitaxial growth of compound semiconductors both for research and production. The factors that have allowed MOCVD to reach this popularity are the purity and abruptness of the grown layers together with the flexibility of the technique, which makes the growth of almost all compound semiconductors possible. This abrupt transition in the composition of the epitaxial structure is necessary for the fabrication of digital or analog alloy system.

The development history of MOCVD technique is equivalent to that of source precursors. Since ZnSe epilayers were grown by MOCVD [16.52], many source precursors of II–VI elements have been developed. $\text{Zn}(\text{CH}_3)_2$ dimethylzinc (DMZn) and $\text{Zn}(\text{C}_2\text{H}_5)_2$ diethylzinc (DEZ) were used at the beginning of MOCVD growth [16.53]. ZnSe and ZnS films were grown using these metalorganic sources and inorganic H_2Se or H_2S . Unfortunately, the quality of these films was very poor. For this reason, $\text{Se}(\text{CH}_3)_2$ dimethylselenide (DMS_{Se}) and $\text{Se}(\text{C}_2\text{H}_5)_2$ diethylselenium (DESe) were developed [16.54]. The quality of epilayers was greatly improved. From then, II and VI elemental gas precursors were proposed one after another. Wright et al. grew a ZnSe film using $(\text{DMZn}(\text{NET}_3)_2)$ triethylamine adduct of dimethylzinc [16.55]. Hirata et al. [16.56] and Nishimura et al. [16.57] proposed methylselenol (MSeH) and tertiarybutylselenol (*t*-BuSeH) as Se sources, respectively. Methylallylselenide (MASE) [16.58], diallyl-selenide (DASE) [16.59], *t*-butylallylselenide (*t*-BuASE) [16.60], tertiarybutyl-selenide (*Dr*-BuSe) [16.61], were also used as Se sources. Fujita et al. found methylmercaptan (MSH) as an S source [16.53]. Besides these, many other source precursors of II–VI elements, such as $\text{Cd}(\text{CH}_3)_2$ dimethylcadmium (DMCd) [16.62], $\text{Te}(\text{CH}_3)_2$ dimethyltelluride (DMTe) [16.62], $\text{Te}(\text{C}_2\text{H}_5)_2$ diethyltelluride (DETe), $\text{Te}(\text{C}_3\text{H}_7)_2$ diisopropyltelluride (DIPTe) [16.63], $\text{S}(\text{C}_2\text{H}_5)_2$ diethylsulfide (DES), $\text{S}(\text{C}_4\text{H}_9)_2$ ditertiarybutylsulfide (DTBS), $(\text{C}_4\text{H}_9)\text{SH}$ tertiarybutylthiol (*t*BuSH), have been used.

The great advantage of using metalorganics is that they are volatile at moderately low temperatures. Since all constituents are in the vapor phase, precise electronic control of gas flow rates and partial pressures is possible. This, combined with pyrolysis reactions that are

relatively insensitive to temperature, allows efficient and reproducible deposition.

The substrate wafer is placed on a graphite susceptor inside a reaction vessel and heated by a radio-frequency (RF) induction heater. The growth temperature depends on the type of compounds grown. Growth is carried out in a hydrogen atmosphere at a pressure of 100–700 torr. The growth precursors decompose on contact with the hot substrate to form epitaxial layers. Each layer is formed by switching the source gases to yield the desired structure.

The films of almost all wide-bandgap II–VI compounds have been grown by MOCVD technique. Most work has been done on p-ZnSe epilayers in the past two decades [16.64–66]. The highest hole concentration of $8.8 \times 10^{17} \text{ cm}^{-3}$ was reported with a NH_3 doping source [16.67]. Recently, quantum wells (QW) and quantum dots (QD) of these wide-bandgap compounds have become the focus. Successful pulsed laser operation at 77 K in ZnCdSe/ZnSe/ZnMgSSe QW-structure separated-confinement heterostructures has been realized [16.68].

MBE and MOMBE

MBE was developed at the beginning of the 1970s to grow high-purity high-quality compound semiconductor epitaxial layers on some substrates [16.69, 70]. To date, it has become a very important technique for growing almost all semiconductor epilayers. An MBE system is basically a vacuum evaporation apparatus. The pressure in the chamber is commonly kept below $\approx 10^{-11}$ torr. Any MBE process is dependent on the relation between the equilibrium vapor pressure of the constituent elements and that of the compound [16.71]. There are a number of features of MBE that are generally considered advantageous for growing semiconducting films: the growth temperature is relatively low, which minimizes any undesirable thermally activated processes such as diffusion; the epilayer thickness can be controlled precisely; and the introduction of different vapor species to modify the alloy composition and to control the dopant concentration can be conveniently achieved by adding different beam cells with proper shutters. These features become particularly important in making structures involving junctions.

Metalorganic molecular-beam epitaxy growth (MOMBE) is one of the variations of the MBE system [16.72, 73]. The difference is that metalorganic gaseous sources are used as the source materials. Therefore, this growth technique has the merits of MOCVD and MBE.

MBE or MOMBE techniques have been used to grow epilayers of almost all wide-bandgap II–VI semiconductors [16.74, 75]. Due to its features, it is very successful in growing super-thin layers, such as single quantum wells (SQW), multiple quantum wells (MQW) [16.76, 77] and nanostructures [16.78].

In nanostructures, quantum dot (QD) structures have attracted a lot of attention in recent years. This field represents one of the most rapidly developing areas of current semiconductor. They present the utmost challenge to semiconductor technology, rendering possible fascinating novel devices. QD are nanometer-size semiconductor structures where charge carriers are confined in all three spatial dimensions. They are neither atomic nor bulk semiconductor, but may best be described as artificial atoms.

In the case of heteroepitaxial growth there are three different growth modes [16.79]: (a) Frank–van der Merwe (FM) or layer-by-layer growth, (b) Volmer–Weber (VW) or island growth, and (c) Stranski–Krastanov (SK) or layer-plus-island growth. Which growth mode will be adopted in a given system depends on the surface free energy of the substrate, (σ_s), that of the film, (σ_f), and the interfacial energy (σ_i). Layer-by-layer growth mode occurs when $\Delta\sigma = \sigma_f + \sigma_i - \sigma_s = 0$. The condition for FM-mode growth is rigorously fulfilled only for homoepitaxy, where $\sigma_s = \sigma_f$ and $\sigma_i = 0$. If the FM-mode growth condition is not fulfilled, then three-dimensional crystals form immediately on the substrate (VW mode). For a system with $\Delta\sigma = 0$ but with a large lattice mismatch between the substrate and the film, initial growth is layer-by-layer. However, the film is strained. As the film grows, the stored strain energy increases. This strained epilayer system can lower its total energy by forming isolated thick islands in which the strain is relaxed by interfacial misfit dislocations, which leads to SK growth in these strained systems. The SK growth mode occurs when there is a lattice mismatch between the substrate and the epilayer, causing the epilayer to be strained, which results in the growth of dot-like self-assembled islands. Wire-like islands can grow from dot-like islands via a shape transition which helps strain relaxation.

For nanostructure fabrication, a thin epilayer is usually grown on a substrate. This two-dimensional (2-D) layer is used to fabricate lower-dimensional structures such as wires (1-D) or dots (0-D) by lithographic techniques. However, structures smaller than the limits of conventional lithography techniques can only be obtained by self-assembled growth utilizing the principles of SK or VW growth. For appropriate growth condi-

tions, self-assembled epitaxial islands can be grown in reasonably well-controlled sizes [16.80].

Because wide-bandgap II–VI materials typically have stronger exciton–phonon interactions than III–V materials, their nanostructures are expected to be very useful in fabricating optoelectronics devices and in exploring the exciton nature in low-dimensional structure. Self-assembled semiconductor nanostructures of different system, such as CdSe/ZnSe [16.81], ZnSe/ZnS [16.82], CdTe/ZnTe [16.83], CdS/ZnSe [16.84], are thought to be advantageous for future application. MBE/MOMBE [16.81, 84], MOCVD [16.82], HWE [16.85] are the main growth techniques used to obtain such structures. MBE is the most advanced technique for the growth of controlled epitaxial layers. With the advancement of nanoscience and nanotechnology, lower-dimensional nanostructures are being fabricated by lithographic techniques from two-dimensional epitaxial layers. Alternately, self-assembled, lower-dimensional nanostructures can be fabricated directly by self-assembly during MBE growth.

Atomic–Layer Epitaxy

ALE is a chemical vapor deposition technique [16.5] where the precise control of the system parameters (pres-

sure and temperature) causes the reaction of adsorption of the precursors to be self-limiting and to stop with the completion of a single atomic layer. The precursors are usually metalorganic molecules. The special feature of ALE is that the layer thickness per cycle is independent of subtle variations of the growth parameters. The growth rate is only dependent on the number of growth cycles and the lattice constant of the deposited material. The conditions for thickness uniformity are fulfilled when material flux on each surface unit is sufficient for monolayer saturation. In an ALE reactor, this means freedom in designing the precursor transport and its interaction with the substrates.

The advantages obtainable with ALE depend on the material to be processed and the type of application. In single-crystal epitaxy, ALE may be a way to obtain a lower epitaxial crystal-growth temperature. It is also a method for making precise interfaces and material layers needed in superlattice structures and super-alloys. In thin-film applications, ALE allows excellent thickness uniformity over large areas. The process has primarily been developed for processing of compound materials. ALE is not only used to grow conventional thin films of II–VI wide-bandgap compounds [16.5,86,87], but is also a powerful method for the preparation of monolayers (ML) [16.88].

16.3 Bulk Crystal Growth

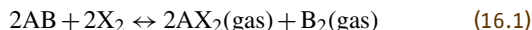
Bulk crystal is the most important subject studied in recent decades. The quality of bulk crystals is the most important aspect of electronic device design. To date, many growth methods have been developed to grow high-quality crystals. Significant improvements have been made in bulk crystal growth with regard to uniformity, reproducibility, thermal stability, diameter control, and impurity and dopant control. According to the phase balance, crystals can be grown from vapor phase, liquid (melt) phase, and solid phase.

16.3.1 The CVT and PVT Techniques

Crystal growth from the the vapor phase is the most basic method. It has advantages that growth can be performed at lower temperatures. This can prevent from phase transition and undesirable contamination. Therefore, this method has commonly been used to grow II–VI compound semiconductors.

Crystal growth techniques from the vapor phase can be divided into chemical vapor transport (CVT) and

physical vapor transport (PVT). CVT is based on chemical transport reactions that occur in a closed ampoule having two different temperature zones. Figure 16.6 shows a typical schematic diagram of the CVT technique. In the high-temperature region, the source AB reacts with the transport agent X:



In the low-temperature region, the reverse reaction takes place. The whole process continues by back-diffusion of the X_2 generated in the lower-temperature region. The transport agent X usually employed is hydrogen (H_2), a halogen (I_2 , Br_2 , Cl_2), a halide (HCl , HBr), and so on. For example, I_2 has been used as a transport agent for ZnS, ZnSe, ZnTe and CdS [16.89]; HCl , H_2 , Cl_2 , NH_3 [16.90], and C and CH_4 [16.91] have been used as the transport agents for ZnO. According to [16.89]: the typical growth temperature for ZnS is 1073–1173 K, for ZnSe 1023–1073 K, for ZnTe 973–1073 K; ΔT is 5–50 K; the concentration of the transport agent is

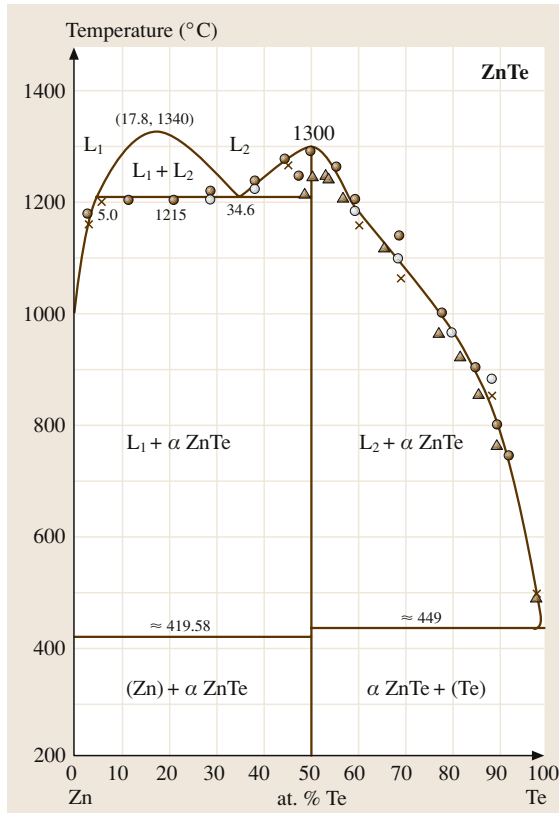
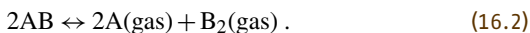


Fig. 16.6 Diagram of a conventional chemical vapor transport system

0.5–5 mg/cm³ of the ampoules vapor space; the aspect ratios are 5–17 at ampoule diameters of 10–20 mm. According to [16.91]; the growth temperature for ZnO is 1228–1273 K, and ΔT is 5–10 K. The transport rate does not strongly depend on the initial amount of carbon when the concentration of the transport agent is over 0.3 mg/cm³.

The PVT method is similar to CVT, but the transport agent is not used. This technique is based on the dissociative sublimation of compounds. Initially, the Piper–Polich method was developed, in 1961 [16.92]. Prior [16.93] improved the PVT method using a reservoir to control the deviation from stoichiometry; the experimental arrangement is shown in Fig. 16.7. The constituent element is placed in the reservoir. The reservoir temperature can be calculated according to the solid–vapor equilibrium



The total pressure (p) in ampoule is given by

$$\begin{aligned} p &= p_A + p_{B_2} \\ &= p_A + K \cdot p_A^{-2} \\ &= (K/p_{B_2})^{1/2} + p_{B_2} , \end{aligned} \quad (16.3)$$

where p_A and p_{B_2} are the partial pressures of the group II and VI elements respectively and $K = p_A^{-2} \cdot p_{B_2}$ is the equilibrium constant of (16.2). At any temperature, there is minimum total pressure (p_{\min}), which corresponds to the condition,

$$p_A = 2p_{B_2} = 2^{1/3} K^{2/3} . \quad (16.4)$$

Under this condition, the vapor-phase composition is stoichiometric and growth rate is maximum [16.94]. After modifying this method to use a closed ampoule, it was applied to grow high-purity and high-quality crystals of II–VI compounds.

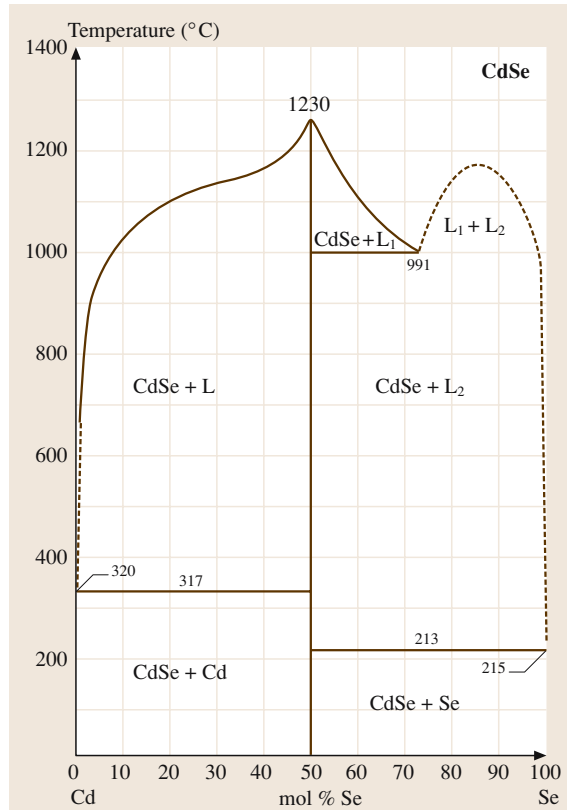


Fig. 16.7 Diagram of the physical vapor transport Piper–Polich method

The **PVT** of II–VI compounds takes advantage of the volatility of both components of the compound semiconductor. This same volatility, coupled with typically high melting points, makes melt growth of these materials difficult. In the **PVT** process, an ampoule containing a polycrystalline source of the desired II–VI compound is heated to a temperature that causes the compound to sublime at a rate conducive to crystal growth. The ampoule is typically placed in a furnace having a temperature gradient over the length of the ampoule, so that the polycrystalline source materials sublime at the end with the higher temperature. The end of the ampoule where the crystal is to be grown is then maintained at a lower temperature. This temperature difference causes supersaturation, and vaporized molecules from source materials eventually deposit at the cooler end. In order to control the deviation from stoichiometry, a reservoir is often used (Fig. 16.7). One of the constituent elements is placed in it. By selecting the proper growth conditions, the rate of deposition can be set to a value leading to growth of high-quality crystals. Typically, **PVT** growth of II–VI compounds is carried out at temperatures much lower than their melting points; this gives benefits in terms of reduced defects, which are related to the melt growth of II–VI compounds such as voids and/or inclusions of excess components of the compound, and also helps to reduce the contamination of the growing crystal from the ampoule. Other effects, such as the reduction of point defects, are also typically found when crystals grown by **PVT** are compared to crystals grown by melt techniques. Although claims have been made that the lower temperatures of physical vapor transport crystal growth should also reduce the twinning found in most of the cubic II–VI compound crystals, the reduction is not usually realized in practice. The assumption that the twinning is a result of cubic/hexagonal phase transitions is not found to be the determining factor in twin formation.

Ohno et al. [16.95] grew cubic ZnS single crystals by the iodine transport method without a seed. By means of Zn-dip treatment, this low-resistivity crystal was used for homoepitaxial **MOCVD** growth, and a metal–insulator–semiconductor(MIS)-structured blue **LED**, which yielded an external quantum efficiency as high as 0.05%. They found that crystal quality was significantly improved by prebaking the ZnS powder in H₂S gas prior to growth. The growth rate also increased by three times.

Isshiki et al. [16.96] purified zinc by a process consisting of vacuum distillation and overlap zone melt-

ing in pure argon. Using refined zinc and commercial high-purity Se, high-quality ZnSe single crystals were grown by the same method, as reported by *Huang and Igaki* [16.97]. The emission intensities of donor-bound exciton (I₂) are remarkably small. The emission intensities of the radiative recombinations of free excitons (E_X) are very strong [16.98]. These intensities indicated the crystal had a very high purity and a very low donor concentration, and they suggest that the purity of the grown crystal strongly depends on the purity of the starting materials. This method is suitable for preparing high-purity crystals, since a purification effect is expected during growth. Impurities with a higher vapor pressure will condense at the reservoir portion and those with a lower vapor pressure will remain in the source crystal. This effect was confirmed by the **PL** results [16.99]. As for these crystals, photoexcited cyclotron resonance measurements have been attempted and cyclotron resonance signals due to electrons [16.100] and heavy holes [16.101] have been detected for the first time. The cyclotron mobility of electrons under $B = 7$ T is 2.3×10^5 cm²/Vs. This indicates that the quality of the grown crystals is very high. Furthermore, the donor concentration in the crystal is estimated to be 4×10^{14} cm³ by analyzing the temperature dependence of the cyclotron mobility [16.99].

The crystals are grown in a self-seeded approach by the **CVT** or **PVT** techniques introduced above. This limits single-crystal volume to several cm³. Meanwhile, grain boundaries and twins are easy to form during growth. In order to solve these problems, seeded chemical vapor transport (**SCVT**) and seeded physical vapor transport (**SPVT**), the so-called modified Lely method, have been developed [16.102]. The difference between **SCVT/SPVT** and **CVT/PVT** is that a seed is set in the crystal growth space before growth starts. The most successful method of eliminating twin formation has usually been by using a polycrystal or single-crystal seed. Even this seeding cannot assure complete elimination of twinning unless seeding is done carefully. The usual method of using small seeds and increasing the diameter of the growing crystal are dependent on the preparation and condition of the walls of the ampoule and the furnace profiles required to eliminate spurious nucleation from the walls. Since the use of a seed crystal provides better control over the nucleation process, high-quality single crystals can be grown [16.103, 104]. Using this technique, sizable single crystals of II–VI wide-bandgap compounds has been commercialized.

Fujita et al. [16.105] grew ZnS single crystals as large as 24 mm × 14 mm × 14 mm by the **SCVT** method

using iodine as a transport agent. The average linear growth rate was about 10^{-6} cm/s. The crystal size depended strongly on the ampoule geometry and the temperature difference between the seed and solvent. The study of the electrical properties showed that the annealed crystal was n-type.

16.3.2 Hydrothermal Growth

The hydrothermal technique is a method for growing crystal from aqueous solvent [16.106]. Figure 16.8 shows a diagram of hydrothermal techniques. The hydrothermal method of crystal growth has several advantages: (1) due to the use of a closed system, it is easier to control oxidization or maintain conditions that allow the synthesis of phases that are difficult to attain by other methods, such as compounds of elements in oxidation states, especially for transition-metal compounds; (2) crystal grow occurs under lower thermal strain, and thus may contain a lower dislocation density than when

the crystal is grown from a melt, where large thermal gradients exist; (3) the method has proven to be very useful for the synthesis of the so-called low-temperature phases; (4) it can be employed for large-scale synthesis of piezoelectric, magnetic, optic, ceramic, and many other special materials; (5) hydrothermal synthesis results in rapid convection and very efficient solute transfer, which results in comparatively rapid growth of larger, purer, and dislocation-free crystals. The most successful example of obtaining II–VI compounds is growing single ZnO crystals.

ZnO crystals are considered extrinsic n-type piezoelectric semiconductors. Undoped crystals have a typical resistivity of $0.1\text{--}100\ \Omega\text{ cm}$ and a drift mobility of $10\text{--}125\text{ cm}^2/\text{Vs}$. Low carrier concentrations can be approached by special growth and annealing methods. ZnO crystal is quite transparent in the range $0.4\text{--}6.0\ \mu\text{m}$. Slight absorption is sometimes found around $2.2\text{--}2.3\ \mu\text{m}$ and an additional slight absorption is found at $3.42\ \mu\text{m}$. Since 1953, Walker [16.107] and

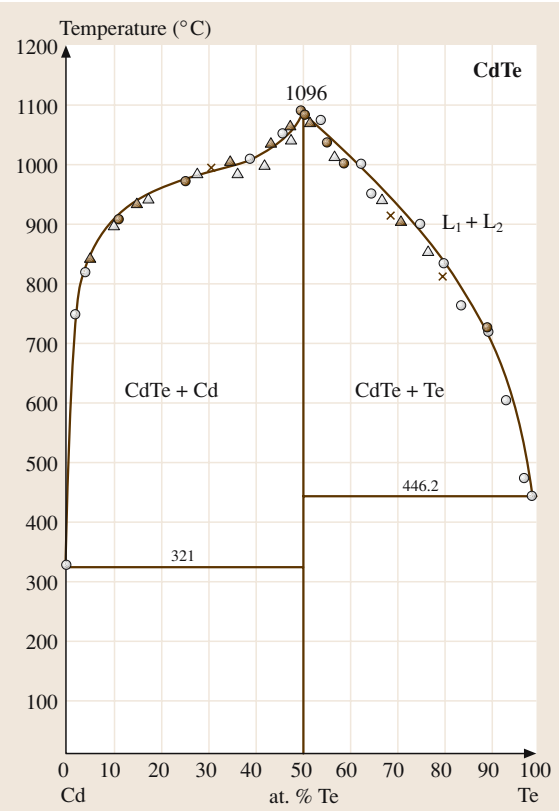


Fig. 16.8 Diagram of a typical hydrothermal technique for growing ZnO single crystals

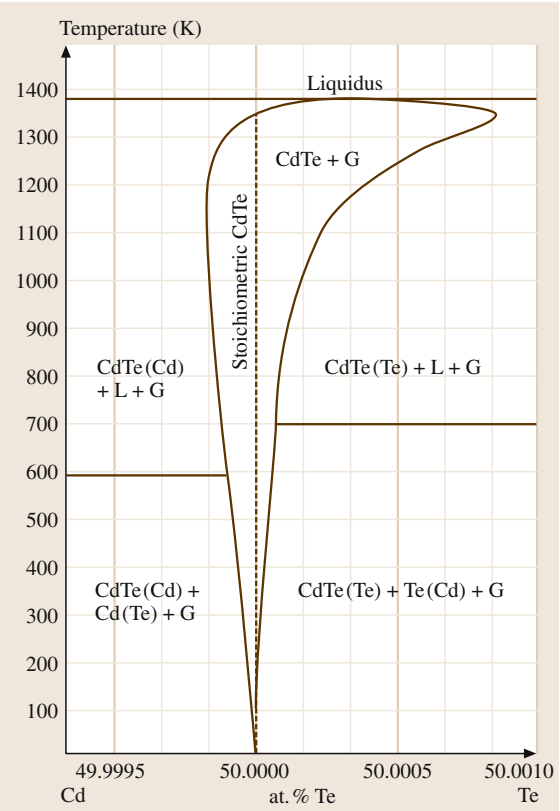


Fig. 16.9a,b Scheme of the vertical Bridgman growth system (a), and its temperature profiles (b)

many other researchers [16.108–111] have grown large ZnO single crystals with hydrothermal techniques and other methods.

The seed, suspended by a Pt wire, and sintered ZnO strings as a source material (nutrient), together with a KOH (3M) and LiOH (1M) aqueous solution, were put into a Pt crucible [16.112] (the hydrothermal conditions are different in different papers). The seed crystals and the source material were separated by a Pt baffle. The crucible was sealed by welding and put into an autoclave. This hydrothermal autoclave is made of high-strength steel. Then, the autoclave was put into a vertical furnace. The temperature of the autoclave was raised to about 673 K, which produced 0.1 GPa of pressure. the growth temperature was monitored by a thermocouple inserted in the autoclave. Seed crystals grew to about 10 mm or bigger after two weeks. The crystal habit of hydrothermal ZnO crystals grown on basal plane seeds shows that growth direction in [0001] is faster 3 times than [000 $\bar{1}$] [16.113].

16.3.3 Bridgman and Gradient Freezing (GF) Method

From the viewpoint of industrial production, melt growth is the most useful for obtaining large single crystals. VPE growth has limitations with regard to crystal size and productivity. The Bridgman technique is a typical crystal growth method from melt. Bridgman growth can be simply understood in terms of a molten charge that passes through a temperature gradient at

a slow speed and solidifies when the temperature is below the melting point of this material. If the ampoule and furnace are stationary and the temperature is gradually reduced by keeping the temperature gradient at the interface constant, this growth process is called the gradient freezing (GF) method [16.114]. In Bridgman or GF growth, single crystals can be grown using either seeded or unseeded ampoules or crucibles.

The Bridgman method has been most extensively used to grow II–VI wide-bandgap compounds such as CdTe, ZnTe and ZnSe crystals, because of the simplicity of the growth apparatus, the high growth rate and the availability of crystals of appropriate size and quality. There are two Bridgman method techniques: the high-pressure technique [16.115] and the closed technique [16.116]. In the former, it is inevitable that a compositional deviation from stoichiometry occurs during melting. Since the properties and structural perfection of these compound crystals are correlated very strongly with this nonstoichiometry [16.117], compositional deviation must be controlled during melt growth. *Omino et al.* [16.116] and *Wang et al.* [16.118, 119] have adopted a closed double crucible to prevent deviation of the melt from stoichiometric composition during Bridgman growth. Figure 16.9 shows a closed vertical Bridgman growth furnace and its temperature profiles.

In Bridgman growth, the temperature gradient (G) and growth rate (R) are very important parameters since they determine the shape of the solid–liquid interface. For this reason, the relationship between temperature gradient and growth rate is investigated. The experi-

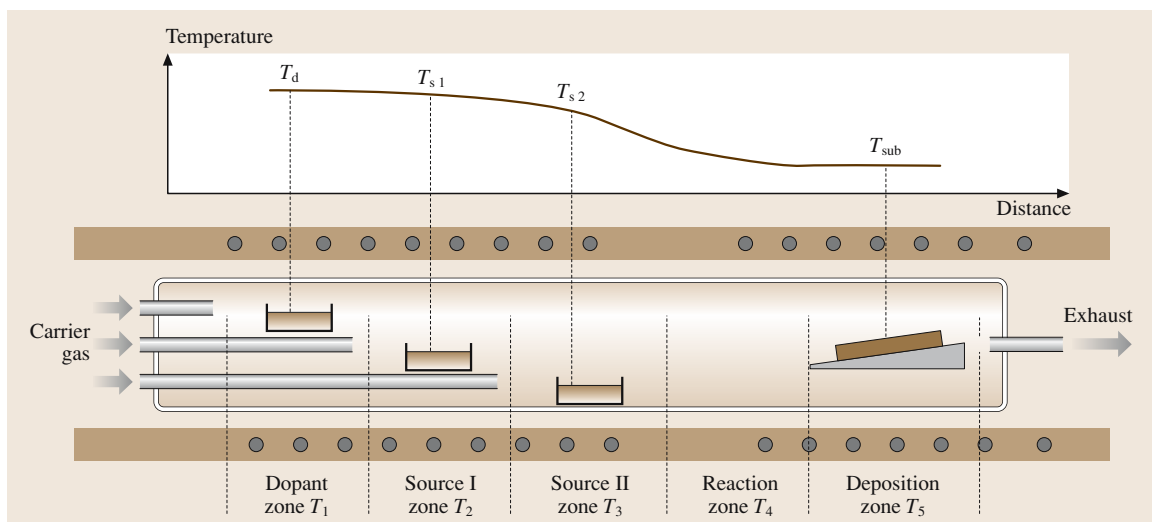


Fig. 16.10 Relationship between growth velocity and temperature gradient at the growth interfaces

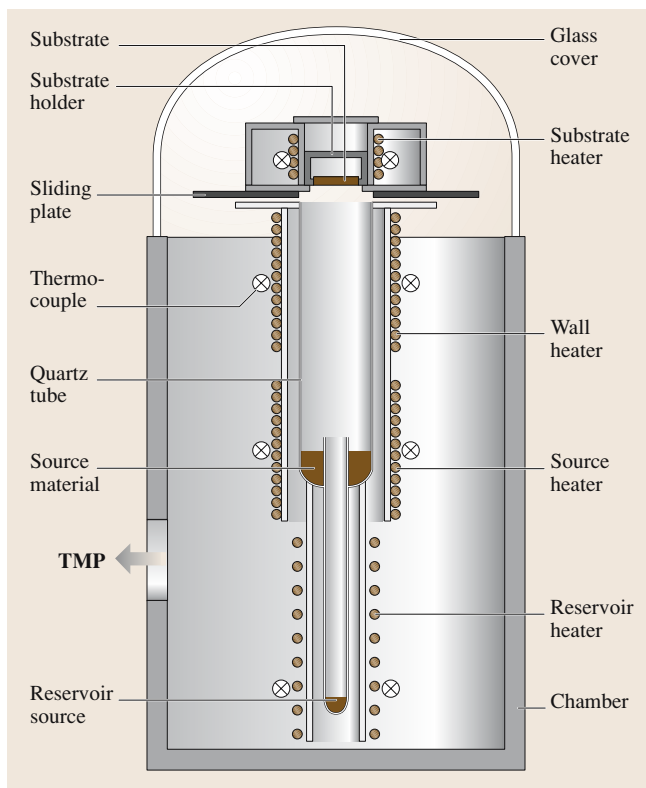


Fig. 16.11 Photograph of twin-free ZnSe single crystal grown by the vertical Bridgman technique

mental results are summarized in Fig. 16.10 [16.120, 121]. The experimental results show that, to grow a ZnSe single crystal, it is necessary that the G/R [(temperature) gradient/(growth) rate] value should be limited to $57\text{--}175\text{ K h/cm}^2$. The most suitable value of G/R , assessed from the determined optimum temperature gradient and growth rate, is 83 K h/cm^2 . Wang et al. [16.118, 119] found the optimum conditions to include: a special temperature program for removing the gas bubbles generated in melt, an overheating temperature of 76 K from the melting point of 1797 K , a temperature gradient of 30 K/cm and a growth rate of 3.6 mm/h as marked by the open squares in Fig. 16.10. Under these growth conditions, twin-free high-quality ZnSe single crystals (Fig. 16.11) were grown using a polycrystalline seed. Chemical etching on the cleaved (110) plane revealed that the average value of the etch pit density (EPD) is about $2 \times 10^5\text{ cm}^{-2}$. The rocking curves of four-crystal XRD showed a full-width at half-maximum (FWHM) value of 19 arcs. The resolved intensive free-exciton, bound-exciton emission

lines and the weak donor-acceptor-pair (DAP) emission bands are observed in the PL spectra at 4.2 K . The FWHM of the I_1^d emission was smaller than 0.5 meV . On the other hand, the deep-level emission bands were almost not observed. All these results suggest that the ZnSe single crystals grown by this method are of very high quality.

Asahi et al. [16.122] successfully grew ZnTe single crystals with a diameter of 80 mm and a length of 50 mm by the vertical gradient freezing (VGF) method. In this method, a high-pressure furnace was used and the melt was encapsulated by B_2O_3 during crystal growth. The growth direction was nearly $\langle 111 \rangle$ or $\langle 110 \rangle$. When long ZnTe crystals were grown, polycrystals were found at the tail. It seems to be difficult to grow an ingot longer than 50 mm . The researchers believed that this is because the shape of the solid-liquid interface easily becomes concave against the liquid at the tail owing to the low thermal conductivity of ZnTe. Evaluation of the crystals showed that the FWHMs of the rocking curve measured by XRD were about 20 arcs, and the EPDs were $5 \times 10^3\text{--}1 \times 10^4\text{ cm}^{-2}$.

16.3.4 The Traveling Heater Method (THM)

The traveling heater method (THM) [16.123] is a solution growth process whereby polycrystalline feed material with an average constant composition is progressively dissolved under the influence of a temperature gradient, followed by deposition in single-crystal form onto a seed with the same composition. The growth proceeds by the relative translation of the heater and charge. THM is particularly useful for the growth of binary and ternary semiconductor alloys, such as CdTe and CdZnTe [16.124]. In such materials, the wide separation between the solidus and liquidus in the pseudo-binary (CdTe–ZnTe) phase diagram imposes a monotonic variation in composition of the solid in the melt growth processes. The THM process ensures a constant macroscale composition in the crystal grown. Since the process takes place at a temperature below the melting point, contamination from the container is also reduced. The reduced operating temperature also leads to a lower ambient pressure within the growth environment, and a reduced risk of ampoule fracture.

16.3.5 Other Methods

Besides the growth methods described, there are many other techniques for growing single crystals of II–VI

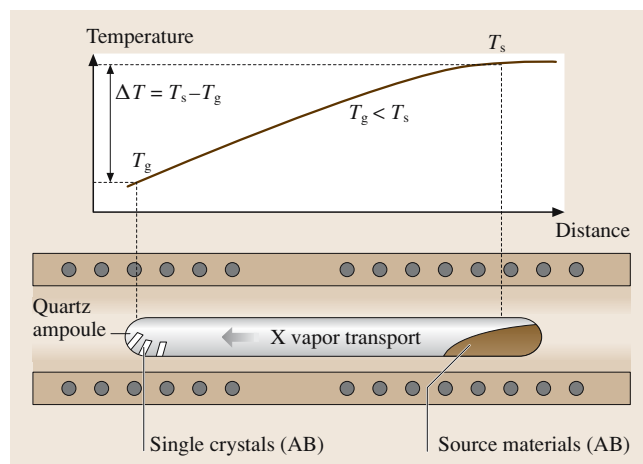
Table 16.3 Growth methods for films and bulk crystals of wide-bandgap II–VI compounds

	Epitaxial growth						Bulk growth			
	LPE	VPE	HWE	ALE	MOCVD	MBE	CVT	PVT	Hydrothermal	Bridgman
ZnS	•	•		•	•	•	•	•		
ZnO					•	•	•		•	
ZnSe	•	•	•	•	•	•	•	•		•
ZnTe	•	•	•	•	•	•	•	•		•
CdS		•	•	•	•	•	•	•		•
CdSe		•	•	•	•	•	•	•		•
CdTe		•	•	•	•	•	•	•		•

wide-bandgap compounds. Zone melting [16.125] and solid-state recrystallization (SSR) [16.126] are often used to grow bulk crystals of ZnSe, ZnS and CdTe.

Recently, *Asahi* et al. [16.127] proved that B_2O_3 is a suitable encapsulant for ZnTe melt growth. Furthermore, B_2O_3 and a total weight of 6N Zn and Te was charged into a pBN crucible. Then this crucible was put into a high-pressure furnace with five heaters. A ZnTe seed was used to pull the ZnTe crystal. Before growth started, the starting materials were heated to 1573 K and kept at this temperature for several hours. The pressure in the growth furnace was kept at 1.5–2 MPa using Ar gas during growth. The temperature gradient on the solid–liquid surface was about 10–20 K/cm. The growth rate was 2–4 mm/h. Under these growth conditions, ZnTe single crystals with a diameter of 80 mm and a height of 40 mm were successfully grown using a combined GF/Kyropoulos method [16.127] (Fig. 16.12).

Some growth methods for film and bulk crystals are summarized in Table 16.3.

**Fig. 16.12** ZnTe single crystal grown by a combination of the GF and Kyropoulos methods

16.4 Conclusions

It was expected that II–VI wide-bandgap compounds would become applicable to optoelectronic devices, especially LEDs and LDs in the short-wavelength visible-light region. However, as they are very strongly bonded with a nearly equal balance of covalent and ionic bonding, it is challenging to grow high-purity high-quality single crystals. Another problem with their application to devices is the difficulty of controlling the conductive type. This is because native defects commonly occur in these semiconductors. These native defects can have either a donor or acceptor character, or even be amphoteric, and they act as compensating

centers. Furthermore, these defects react with dopant impurities to form complexes. This makes it difficult to reverse the conductive type. For example, in the case of ZnSe, it is still difficult to obtain low-resistivity p-type crystals or epilayers that can be used to fabricate devices. Therefore, a significant improvement in the understanding of the fundamental physical and chemical properties needs to be achieved. In particular, the specifications for many applications are very demanding, and considerable progress needs to be made in growth, particularly in the areas of reproducibility, convenient shape, conductivity, and structural perfection.

References

- 16.1 A. Lopez-Otero: *Thin Solid Films* **49**, 1 (1978)
- 16.2 H. M. Manasevit, W. I. Simpson: *J. Electrochem. Soc.* **118**, 644 (1971)
- 16.3 L. L. Chang, R. Ludeke: *Epitaxial Growth, Part A*, ed. by J. W. Matthews (Academic, New York 1975) p. 37
- 16.4 E. Veuhoff, W. Pletschen, P. Balk, H. Luth: *J. Cryst. Growth* **55**, 30 (1981)
- 16.5 T. Suntola: *Mater. Sci. Rep.* **4**, 261 (1989)
- 16.6 M. M. Faktor, R. Heckingbottom, I. Garrett: *J. Cryst. Growth* **9**, 3 (1971)
- 16.7 I. Kikuma, M. Furukoshi: *J. Cryst. Growth* **41**, 103 (1977)
- 16.8 Y. V. Korostelin, V. J. Kozlovskij, A. S. Nasibov, P. V. Shapkin: *J. Cryst. Growth* **159**, 181 (1996)
- 16.9 J. F. Wang, A. Omino, M. Isshiki: *Mater. Sci. Eng.* **83**, 185 (2001)
- 16.10 S. H. Song, J. F. Wang, G. M. Lalev, L. He, M. Isshiki: *J. Cryst. Growth* **252**, 102 (2003)
- 16.11 H. Harmann, R. Mach, B. Sell: In: *Current Topics Mater. Sci.*, Vol. 9, ed. by E. Kaldis (North-Holland, Amsterdam 1982) pp. 1–414
- 16.12 P. Rudolph, N. Schäfer, T. Fukuda: *Mater. Sci. Eng.* **15**, 85 (1995)
- 16.13 R. Shetty, R. Balasubramanian, W. R. Wilcox: *J. Cryst. Growth* **100**, 51 (1990)
- 16.14 K. W. Böer: *Survey of Semiconductor Physics, Vol. 1: Electrons and Other Particles in Bulk Semiconductors* (Van Nostrand, New York 1990)
- 16.15 C. M. Wolf, N. Holonyak, G. E. Stillman: *Physical Properties of Semiconductors* (Prentice Hall, New York 1989)
- 16.16 L. Smart, E. Moore: *Solid State Chemistry*, 2nd edn. (Chapman Hall, New York 1995)
- 16.17 E. Lide (Ed.): *Handbook of Chemistry and Physics*, 2nd edn. (CRC, Boca Raton 1973)
- 16.18 J. Singh: *Physics of Semiconductors and Their Heterostructures* (McGraw-Hill, New York 1993)
- 16.19 N. Yamamoto, H. Horinaka, T. Miyauchi: *Jpn. J. Appl. Phys.* **18**, 225 (1997)
- 16.20 H. Neumann: *Kristall Technik* **15**, 849 (1980)
- 16.21 J. Camassel, D. Auvergne, H. Mathieu: *J. Phys. Colloq.* **35**, C3–67 (1974)
- 16.22 W. Shan, J. J. Song, H. Luo, J. K. Furdyna: *Phys. Rev.* **50**, 8012 (1994)
- 16.23 K. A. Dmitrenko, S. G. Shevel, L. V. Taranenko, A. V. Marintchenko: *Phys. Status Solidi B* **134**, 605 (1986)
- 16.24 S. Logothetidis, M. Cardona, P. Lautenschlager, M. Garriga: *Phys. Rev. B* **34**, 2458 (1986)
- 16.25 R. C. Sharma, Y. A. Chang: *J. Cryst. Growth* **88**, 192 (1988)
- 16.26 H. Okada, T. Kawanaka, S. Ohmoto: *J. Cryst. Growth* **165**, 31 (1996)
- 16.27 N. Kh. Abrikosov, V. F. Bankina, L. B. Poretzkaya, E. V. Skudnova, S. N. Chichevskaya: *Poluprovodnikovye chalkogenidy i splavy na ikh osnovje* (Nauka, Moscow 1975) (in Russian)
- 16.28 R. F. Brebrick: *J. Cryst. Growth* **86**, 39 (1988)
- 16.29 M. R. Lorenz: *Physics and Chemistry of II–VI Compounds*, ed. by M. Aven, J. S. Prener (North Holland, Amsterdam 1967) pp. 210–211
- 16.30 T. Yao: *Optoelectron. Dev. Technol.* **6**, 37 (1991)
- 16.31 H. Nakamura, M. Aoki: *Jpn. J. Appl. Phys.* **20**, 11 (1981)
- 16.32 C. Werkhoven, B. J. Fitzpatrick, S. P. Herko, R. N. Bhargava, P. J. Dean: *Appl. Phys. Lett.* **38**, 540 (1981)
- 16.33 H. Nakamura, S. Kojima, M. Wasgiyama, M. Aoki: *Jpn. J. Appl. Phys.* **23**, L617 (1984)
- 16.34 V. M. Skobeeva, V. V. Serdyuk, L. N. Semenyuk, N. V. Malishin: *J. Appl. Spectrosc.* **44**, 164 (1986)
- 16.35 P. Lilley, P. L. Jones, C. N. W. Litting: *J. Mater. Sci.* **5**, 891 (1970)
- 16.36 T. Matsumoto, T. Morita, T. Ishida: *J. Cryst. Growth* **53**, 225 (1987)
- 16.37 S. Zhang, H. Kinto, T. Yatabe, S. Iida: *J. Cryst. Growth* **86**, 372 (1988)
- 16.38 S. Iida, T. Yatabe, H. Kinto: *Jpn. J. Appl. Phys.* **28**, L535 (1989)
- 16.39 P. Besomi, B. W. Wessels: *J. Cryst. Growth* **55**, 477 (1981)
- 16.40 T. Kyotani, M. Isshiki, K. Masumoto: *J. Electrochem. Soc.* **136**, 2376 (1989)
- 16.41 N. Stucheli, E. Bucher: *J. Electron. Mater.* **18**, 105 (1989)
- 16.42 M. Nishio, Y. Nakamura, H. Ogawa: *Jpn. J. Appl. Phys.* **22**, 1101 (1983)
- 16.43 N. Lovergine, R. Cingolani, A. M. Mancini, M. Ferrara: *J. Cryst. Growth* **118**, 304 (1992)
- 16.44 O. De. Melo, E. Sánchez, S. De. Roux, F. Rábago-Bernal: *Mater. Chem. Phys.*, **59**, 120 (1999)
- 16.45 M. Kasuga, H. Futami, Y. Iba: *J. Cryst. Growth* **115**, 711 (1991)
- 16.46 J. F. Wang, K. Kikuchi, B. H. Koo, Y. Ishikawa, W. Uchida, M. Isshiki: *J. Cryst. Growth* **187**, 373 (1998)
- 16.47 J. Humenberger, G. Linnet, K. Lischka: *Thin Solid Films* **121**, 75 (1984)
- 16.48 F. Sasaki, T. Mishina, Y. Masumoto: *J. Cryst. Growth* **117**, 768 (1992)
- 16.49 B. J. Kim, J. F. Wang, Y. Ishikawa, S. Sato, M. Isshiki: *Phys. Stat. Sol. (a)* **191**, 161 (2002)
- 16.50 A. Rogalski, J. Piotrowski: *Prog. Quantum Electron.* **12**, 87 (1988)
- 16.51 G. M. Lalev, J. Wang, S. Abe, K. Masumoto, M. Isshiki: *J. Crystal Growth* **256**, 20 (2003)
- 16.52 H. M. Manasevit: *Appl. Phys. Lett.* **12**, 1530 (1968)

- 16.53 Sg. Fujita, M. Isemura, T. Sakamoto, N. Yoshimura: J. Cryst. Growth **86**, 263 (1988)
- 16.54 H. Mitsuhashi, I. Mitsuishi, H. Kukimoto: J. Cryst. Growth **77**, 219 (1986)
- 16.55 P. J. Wright, P. J. Parbrook, B. Cockayne, A. C. Jones, E. D. Orrell, K. P. O'Donnell, B. Henderson: J. Cryst. Growth **94**, 441 (1989)
- 16.56 S. Hirata, M. Isemura, Sz. Fujita, Sg. Fujita: J. Cryst. Growth **104**, 521 (1990)
- 16.57 S. Nishimura, N. Iwasa, M. Senoh, T. Mukai: Jpn. J. Appl. Phys. **32**, L425 (1993)
- 16.58 K. P. Giapis, K. F. Jensen, J. E. Potts, S. J. Pachuta: Appl. Phys. Lett. **55**, 463 (1989)
- 16.59 S. J. Pachuta, K. F. Jensen, S. P. Giapis: J. Cryst. Growth **107**, 390 (1991)
- 16.60 M. Daneš, J. S. Huh, L. Foley, K. F. Jensen: J. Cryst. Growth **145**, 530 (1994)
- 16.61 W. Kuhn, A. Naumov, H. Stanzl, S. Bauer, K. Wolf, H. P. Wagner, W. Gebhardt, U. W. Pohl, A. Krost, W. Richter, U. Dümichen, K. H. Thiele: J. Cryst. Growth **123**, 605 (1992)
- 16.62 J. K. Menno, J. W. Kerri, F. H. Robert: J. Phys. Chem. B **101**, 4882 (1997)
- 16.63 H. P. Wagner, W. Kuhn, W. Gebhardt: J. Cryst. Growth **101**, 199 (1990)
- 16.64 N. R. Taskar, B. A. Khan, D. R. Dorman, K. Shahzad: Appl. Phys. Lett. **62**, 270 (1993)
- 16.65 Y. Fujita, T. Terada, T. Suzuki: Jpn. J. Appl. Phys. **34**, L1034 (1995)
- 16.66 J. Wang, T. Miki, A. Omino, K. S. Park, M. Isshiki: J. Cryst. Growth **221**, 393 (2000)
- 16.67 M. K. Lee, M. Y. Yeh, S. J. Guo, H. D. Huang: J. Appl. Phys. **75**, 7821 (1994)
- 16.68 A. Toda, T. Margalith, D. Imanishi, K. Yanashima, A. Ishibashi: Electron. Lett. **31**, 1921 (1995)
- 16.69 A. Cho: J. Vac. Sci. Tech. **8**, S31 (1971)
- 16.70 C. T. Foxon: J. Cryst. Growth **251**, 130 (2003)
- 16.71 T. Yao: *The Technology and Physics of Molecular Beam Epitaxy*, ed. by E. H. C. Parker (Plenum, New York 1985) Chap. 10, p. 313
- 16.72 E. Veuhoff, W. Pletschen, P. Balk, H. Luth: J. Cryst. Growth **55**, 30 (1981)
- 16.73 M. B. Panish, S. Sumski: J. Appl. Phys. **55**, 3571 (1984)
- 16.74 Y. P. Chen, G. Brill, N. K. Dhar: J. Cryst. Growth **252**, 270 (2003)
- 16.75 H. Kato, M. Sano, K. Miyamoto, T. Yao: J. Cryst. Growth **237–239**, 538 (2002)
- 16.76 M. Imaizumi, M. Adachi, Y. Fujii, Y. Hayashi, T. Soga, T. Jimbo, M. Umeno: J. Cryst. Growth **221**, 688 (2000)
- 16.77 W. Xie, D. C. Grillo, M. Kobayashi, R. L. Gunshor, G. C. Hua, N. Otsuka, H. Jeon, J. Ding, A. V. Nurmikko: Appl. Phys. Lett. **60**, 463 (1992)
- 16.78 S. Guha, A. Madhukar, K. C. Rajkumar: Appl. Phys. Lett. **57**, 2110 (1990)
- 16.79 E. Bauer, J. H. van der Merwe: Phys. Rev. B **33**, 3657 (1986)
- 16.80 J. Drucker, S. Chapparro: Appl. Phys. Lett. **71**, 614 (1997)
- 16.81 S. H. Xin, P. D. Wang, A. Yin, C. Kim, M. Dobrowolska, J. L. Merz, J. K. Furdyna: Appl. Phys. Lett. **69**, 3884 (1996)
- 16.82 M. C. Harris Liao, Y. H. Chang, Y. H. Chen, J. W. Hsu, J. M. Lin, W. C. Chou: Appl. Phys. Lett. **70**, 2256 (1997)
- 16.83 Y. Terai, S. Kuroda, K. Takita, T. Okuno, Y. Masumoto: Appl. Phys. Lett. **73**, 3757 (1998)
- 16.84 M. Kobayashi, S. Nakamura, K. Wakao, A. Yoshikawa, K. Takahashi: J. Vac. Sci. Technol. B **16**, 1316 (1998)
- 16.85 S. O. Ferreira, E. C. Paiva, G. N. Fontes, B. R. A. Neves: J. Appl. Phys. **93**, 1195 (2003)
- 16.86 M. A. Herman, J. T. Sadowski: Cryst. Res. Technol. **34**, 153 (1999)
- 16.87 M. Ahonen, M. Pessa, T. Suntola: Thin Solid Films **65**, 301 (1980)
- 16.88 M. Ritala, M. Leskelä: Nanotechnology **10**, 19 (1999)
- 16.89 H. Hartmann: J. Cryst. Growth **42**, 144 (1977)
- 16.90 M. Shiloh, J. Gutman: J. Cryst. Growth **11**, 105 (1971)
- 16.91 S. Hassani, A. Tromson-Carli, A. Lusson, G. Didier, R. Triboulet: Phys. Stat. Sol. (b) **229**, 835 (2002)
- 16.92 W. W. Piper, S. J. Polich: J. Appl. Phys. **32**, 1278 (1961)
- 16.93 A. C. Prior: J. Electrochem. Soc. **108**, 106 (1961)
- 16.94 T. Kiyosawa, K. Igaki, N. Ohashi: Trans. Jpn. Inst. Metals **13**, 248 (1972)
- 16.95 T. Ohno, K. Kurisu, T. Taguchi: J. Cryst. Growth **99**, 737 (1990)
- 16.96 M. Isshiki, T. Tomizono, T. Yoshita, T. Ohkawa, K. Igaki: J. Jpn. Inst. Metals **48**, 1176 (1984)
- 16.97 X. M. Huang, K. Igaki: J. Cryst. Growth **78**, 24 (1986)
- 16.98 M. Isshiki, T. Yoshita, K. Igaki, W. Uchida, S. Suto: J. Cryst. Growth **72**, 162 (1985)
- 16.99 M. Isshiki: J. Cryst. Growth **86**, 615 (1988)
- 16.100 T. Ohyama, E. Otsuka, T. Yoshita, M. Isshiki, K. Igaki: Jpn. J. Appl. Phys. **23**, L382 (1984)
- 16.101 T. Ohyama, K. Sakakibara, E. Otsuka, M. Isshiki, K. Igaki: Phys. Rev. B **37**, 6153 (1988)
- 16.102 Y. M. Tairov, V. F. Tsvetkov: J. Cryst. Growth **43**, 209 (1978)
- 16.103 G. Cantwell, W. C. Harsch, H. L. Cotal, B. G. Markey, S. W. S. McKeever, J. E. Thomas: J. Appl. Phys. **71**, 2931 (1992)
- 16.104 Yu. V. Korostelin, V. I. Kozlovsky, A. S. Nasibov, P. V. Shapkin: J. Cryst. Growth **161**, 51 (1996)
- 16.105 S. Fujita, H. Mimoto, H. Takebe, T. Noguchi: J. Cryst. Growth **47**, 326 (1979)
- 16.106 K. Byrappa: *Hydrothermal Growth of Crystal*, ed. by K. Byrappa (Pergamon, Oxford 1991)
- 16.107 A. C. Walker: J. Am. Ceram. Soc. **36**, 250 (1953)
- 16.108 R. A. Laudice, E. D. Kolg, A. J. Caporaso: J. Am. Ceram. Soc. **47**, 9 (1964)
- 16.109 M. Suscavage, M. Harris, D. Bliss, P. Yip, S.-Q. Wang, D. Schwall, L. Bouthillette, J. Bailey, M. Callahan, D. C. Look, D. C. Reynolds, R. L. Jones, C. W. Litton: MRS Internet J. Nitride Semicond. Res **4S1**, G3.40 (1999)

- 16.110 L. N. Demianets, D. V. Kostomarov: *Ann. Chim. Sci. Mater.* **26**, 193 (2001)
- 16.111 N. Ohashi, T. Ohgaki, T. Nakata, T. Tsurumi, T. Sekiguchi, H. Haneda, J. Tanaka: *J. Kor. Phys. Soc.* **35**, S287 (1999)
- 16.112 D. C. Look, D. C. Reynolds, J. R. Sizelove, R. L. Jones, C. W. Litton, G. Gantwell, W. C. Harsch: *Solid State Commun.* **105**, 399 (1988)
- 16.113 T. Sekiguchi, S. Miyashita, K. Obara, T. Shishido, N. Sakagami: *J. Cryst. Growth* **214/215**, 72 (2000)
- 16.114 P. Höschl, Yu. M. Ivanov, E. Belas, J. Franc, R. Grill, D. Hlidak, P. Moravec, M. Zvara, H. Sitter, A. Toth: *J. Cryst. Growth* **184/185**, 1039 (1998)
- 16.115 T. Fukuda, K. Umetsu, P. Rudolph, H. J. Koh, S. Iida, H. Uchiki, N. Tsuboi: *J. Cryst. Growth* **161**, 45 (1996)
- 16.116 A. Omino, T. Suzuki: *J. Cryst. Growth* **117**, 80 (1992)
- 16.117 I. Kikuma, M. Furukoshi: *J. Cryst. Growth* **71**, 136 (1985)
- 16.118 J. F. Wang, A. Omino, M. Isshiki: *J. Cryst. Growth* **214/215**, 875 (2000)
- 16.119 J. Wang, A. Omino, M. Isshiki: *J. Cryst. Growth* **229**, 69 (2001)
- 16.120 J. F. Wang, A. Omino, M. Isshiki: *Mater. Sci. Eng. B* **83**, 185 (2001)
- 16.121 P. Rudolph, N. Schäfer, T. Fukuda: *Mater. Sci. Eng. R* **15**, 85 (1995)
- 16.122 T. Asahi, A. Arakawa, K. Sato: *J. Cryst. Growth* **229**, 74 (2001)
- 16.123 M. Ohmori, Y. Iwase, R. Ohno: *Mater. Sci. Eng. B* **16**, 283 (1999)
- 16.124 R. Triboulet: *Prog. Cryst. Growth Char. Mater.* **128**, 85 (1994)
- 16.125 H. H. Woodbury, R. S. Lewandowski: *J. Cryst. Growth* **10**, 6 (1971)
- 16.126 R. Triboulet: *Cryst. Res. Technol.* **38**, 215 (2003)
- 16.127 T. Asahi, T. Yabe, K. Sato: *The Japan Society of Applied Physics and Related Societies, Extended Abstracts, The 50th Spring Meeting, (2003) p. 332*

Structural Ch

17. Structural Characterization

The aim of this chapter is to convey the basic principles of X-ray and electron diffraction, as used in the structural characterization of semiconductor heterostructures. A number of key concepts associated with radiation–material and particle–material interactions are introduced, with emphasis placed on the nature of the signal used for sample interrogation. Various modes of imaging and electron diffraction are then described, followed by a brief appraisal of the main techniques used to prepare electron-transparent membranes for TEM analysis. A number of case studies on electronic and photonic material systems are then presented in the context of a growth or device development program; these emphasize the need to use complementary techniques when characterizing a given heterostructure.

17.1 Radiation–Material Interactions	344
17.2 Particle–Material Interactions	345
17.3 X-Ray Diffraction	348
17.4 Optics, Imaging and Electron Diffraction	351
17.4.1 Electron Diffraction and Image Contrast Analysis	355
17.4.2 Microdiffraction and Polarity	358
17.4.3 Reflection High–Energy Electron Diffraction	359
17.5 Characterizing Functional Activity	362
17.6 Sample Preparation	362
17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials	364
17.7.1 Identifying Defect Sources Within Homoepitaxial GaN	366
17.7.2 Cathodoluminescence/Correlated TEM Investigation of Epitaxial GaN	367
17.7.3 Scanning Transmission Electron Beam Induced Conductivity of Si/Si _{1–x} Ge _x /Si(001)	367
17.8 Concluding Remarks	370
References	370

The functional properties of semiconductors emanate from their atomic structures; indeed, the interrelationship between materials processing, microstructure and functional properties lies at the heart of semiconductor science and technology. Therefore, if we are to elucidate how the functional properties of a semiconductor depend on the processing history (the growth or device fabrication procedures used), then we must study the development of the microstructure of the semiconductor by applying an appropriate combination of analytical techniques to the given bulk crystal, heterostructure or integrated device structure.

The main aim of this chapter is to provide a general introduction to the techniques used to characterize the structures of semiconductors. Thus, we consider techniques such as X-ray diffraction (XRD) and electron diffraction, combined with diffraction contrast imaging, alongside related techniques used for chem-

ical microanalysis, since modern instruments such as analytical electron microscopes (AEMs) provide a variety of operational modes that allow both structure and chemistry to be investigated, in addition to functional activity. For example, chemical microanalyses of the fine-scale structures of materials can be performed within a scanning electron microscope (SEM) and/or a transmission electron microscope (TEM), using the techniques of energy dispersive X-ray (EDX) analysis, wavelength dispersive X-ray (WDX) analysis or electron energy loss spectrometry (EELS). In addition, electrical and optical properties of semiconductors can also be investigated in situ using the techniques of electron beam induced conductivity (EBIC) or cathodoluminescence (CL), respectively. Techniques such as X-ray photoelectron spectrometry (XPS; also known as electron spectroscopy for chemical analysis, ESCA), secondary ion mass spectrometry (SIMS) or Ruther-

ford backscattering spectrometry (RBS), can also be used to study semiconductor chemistry. We should also mention reflection high-energy electron diffraction (RHEED), which can be used for the rapid structural assessment of the near surface of bulk or thin film samples.

A far from exhaustive list of acronyms one might come across when assessing a given sample is incorporated into the general list of abbreviations at the start of this book. We can organize these techniques used to characterize materials into groups based on a number of viewpoints: for example, with respect to the material property being investigated; whether they are destructive or nondestructive; bulk or near-surface assessment techniques; based on radiation–material or particle–material interactions, or based on elastic or inelastic scattering processes, or whether they are diffraction-, imaging- or spectroscopic-based techniques.

In this broad introduction to the structural characterization of semiconductors, we focus on the interaction of

a material with radiation and/or particles. Thus, we start by considering the interactions of photons, electrons or ions with a sample and the nature of the signals used for structural or chemical microanalysis. We then briefly focus on the techniques of X-ray and electron diffraction, and issues regarding the formation of TEM images. The aim is simply to convey an appreciation of the underlying principles, the applicability of these characterization techniques and the information provided by them. We also briefly consider sample preparation, and the chapter closes with a variety of TEM-based case studies of semiconductor heterostructures, which are included to illustrate some of the approaches used to characterize fine-scale microstructure, as well as to emphasise the need for complementary analysis when assessing the interrelationships between processing, structure and functional properties. Much literature already exists in this area, and a selection of references is provided at the end of the chapter that tackle many of the topics we cover here in more detail [17.1–15].

17.1 Radiation–Material Interactions

Each part of the electromagnetic spectrum has quantum energies that can be used to elicit certain forms of excitation at the atomic or molecular level. Different parts of the electromagnetic spectrum will interact with matter in different ways, according to the energy states within the material, allowing absorption or ionization effects to

occur. The salient features of these various radiation–material interaction processes are summarized in the schematics shown in Fig. 17.1.

As the quantum energy increases from radio waves, through microwaves, to infrared and visible light, absorption increases, whilst specific quantized ionization effects come into play upon moving further into the ultraviolet and X-ray parts of the spectrum. Microwave and infrared radiation, for example, can interact with the quantum states of molecular rotation and torsion, leading to the generation of heat for example. Strong absorption also occurs within metallic conductors, leading to the induction of electric currents. Visible and ultraviolet light can elevate electrons to higher energy levels in what is known as the photoelectric effect, which is essentially the liberation of electrons from matter by short-wavelength electromagnetic radiation when all of the incident radiation energy is transferred to an electron. This process can be explained in terms of the absorption of discrete photon energies, with electrons being emitted when the photon energy exceeds the material's work function for the case of weakly bound electrons or the binding energy for more strongly bound inner shell electrons (Fig. 17.2a).

X-ray and γ -ray quantum energies are generally too high to be completely absorbed in direct electron

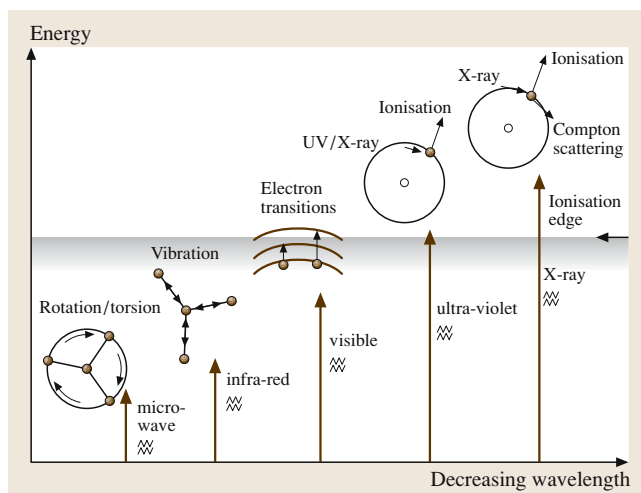


Fig. 17.1 Schematic illustrating the various interactions of energetic radiation with a molecule or atom

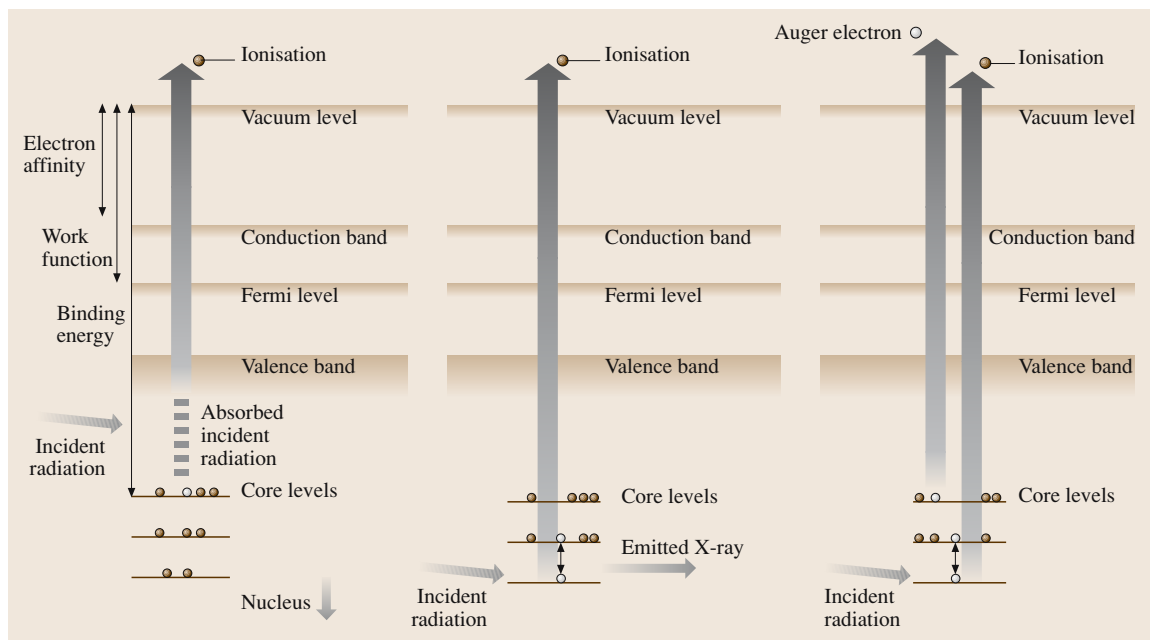


Fig. 17.2a–c Schematics illustrating (a) the photoelectric effect; (b) characteristic X-ray emission and (c) Auger electron emission

transitions, but they can induce ionization, with the displacement of electrons from atoms to form ions. The relaxation of a high-energy electron to the vacated state can lead to the emission of a characteristic X-ray (Fig. 17.2b) or the emission of an Auger electron (Fig. 17.2c). During the process of ionization, some of the incident pho-

ton energy is transferred to the ejected electron in the form of kinetic energy, and the scattering of a lower energy photon (longer wavelength X-ray) occurs, termed Compton scattering. X-rays can also be scattered elastically by shell electrons, without the loss of energy, through a process of absorption and re-emission.

17.2 Particle–Material Interactions

The interaction of an energetic particle with the surface of a material is most commonly associated with the process of sputtering – the non-thermal removal of atoms from a surface under ion bombardment. Transfer of momentum to the surface atoms is followed by a chain of collision events leading to the ejection of matrix atoms. Figure 17.3 illustrates the various signals produced when a particle interacts with a material, depending on the energy available. In broad terms, processes in the range 10^4 – 10^5 eV are associated with ion implantation, the 10 – 10^3 eV energy range is associated with sputtering, whilst the creation of activated point defects occurs in the 1 – 10^3 eV range. At lower 0.1 – 100 eV energies, desorption of surface impurity atoms occurs, whilst energies in the range 0.01 – 1 eV are associated with the enhanced

mobility of surface condensing particles, such as those required during growth. In practical terms, the process of sputtering is most efficient when the masses of the incident and ejected particles are similar, whilst it is also dependent on the sputtering gas pressure, the energy spread of the particles, the bias conditions and the sample geometry.

In this context, we should briefly mention the technique of SIMS, which uses heavy ions, typically in the range 2 – 30 keV, with an ion current density of ≈ 1 mA/cm². The sputtered atoms consist of a mixture of neutrals and ions, and the latter may be mass-spectrally analyzed in order to perform elemental depth profiling. The sensitivity of the technique is very high – for example on the scale of dopant concentrations

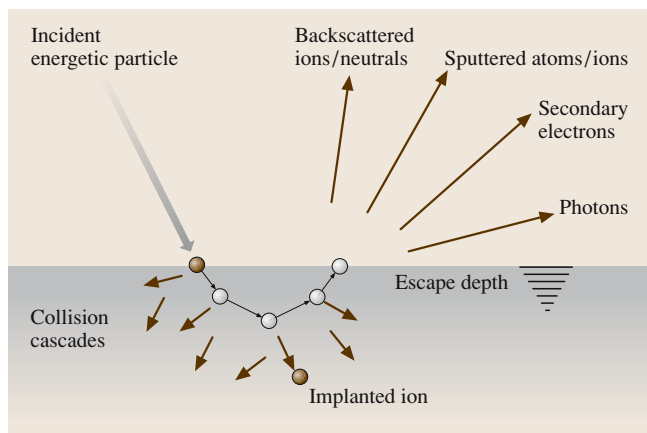


Fig. 17.3 Schematic illustrating the signals produced by the interaction of an energetic ion beam with a sample

within semiconductors – but standards are required for quantitative analysis. One variant of **SIMS** makes use of lower energy primary ions ($0.5\text{--}2\text{ keV}$, $\approx 1\text{ nA/cm}^2$) with an almost negligible sputter rate, thus enabling surface chemical analysis. Conversely, the technique of **RBS** makes use of a very high energy ($2\text{--}3\text{ MeV}$) beam of light ions bombarding a sample normal to its surface. An ion such as helium is chosen to avoid the effects of sputtering, whilst high energy is required to overcome the problem of ion neutralization and the screening interaction potential between the ion and the nucleus associated with techniques such as low-energy ion scattering (**LEIS**). As an energetic positive ion penetrates the sample it loses energy, mainly due to collisions with electrons, but occasionally (and more significantly) with

nuclei. The energy of the backscattered ion depends on the depth of the collision and the mass of the target atom. By measuring the energy spectrum of the recoiling ions, information on elemental composition and depth within a sample can be obtained.

By way of comparison, Fig. 17.4 illustrates the variety of signals produced from the interaction of an energetic beam of electrons with a thin semiconductor sample. In order to make sense of the origins of the many different signals, we must consider the phenomenon of electron scattering, which underpins them all. The process of elastic (Rutherford) scattering arises from an electrostatic (Coulomb) interaction with the nucleus and surrounding electrons of an atom, leading to a change in the incident electron direction without loss of energy. Elastically scattered electrons contribute to the formation of diffraction patterns and diffraction contrast images in **TEM**. Conversely, inelastically scattered electrons have, by definition, lost a certain amount of energy. In this context, core-shell interaction processes produce scattered electrons whose energies depend on the atomic number of the scatterer, and an analysis of the loss of electron energy (up to $\approx 1000\text{ eV}$) is the basis of the technique of electron energy loss spectroscopy (**EELS**). In addition, plasmon scattering ($5\text{--}30\text{ eV}$ loss) can occur, due to the interactions of incident electrons with waves in the conduction band of a metal, or the bonding electrons of non-metals. The signatures from plasmon-scattered electrons can dominate the low-energy regimes of **EELS** spectra, providing information on sample thickness. Phonon scattering ($\approx 1\text{ eV}$ loss) can also occur, which is the interaction of incident electrons with quantized atomic vibrations within a sample, leading to the production of heat. In an interaction of an electron beam with a bulk sample, nearly all of the incident energy ends up being dissipated through such phonon interactions. The probability of each type of scattering interaction is commonly expressed either as a cross-section, representing the apparent area the scattering process presents to the electron, or as a mean free path, which is the average distance the electron travels before being scattered.

A variety of secondary events also occur as a direct consequence of these primary electron scattering processes. For example, following interactions with a high-energy beam of electrons, excited atoms may subsequently relax in a number of different ways (in a similar fashion to the processes induced by incident high-energy X-ray photons). If core-level (inner shell) electrons are displaced, the relaxation of electrons from higher energy shells to the lower energy core states can

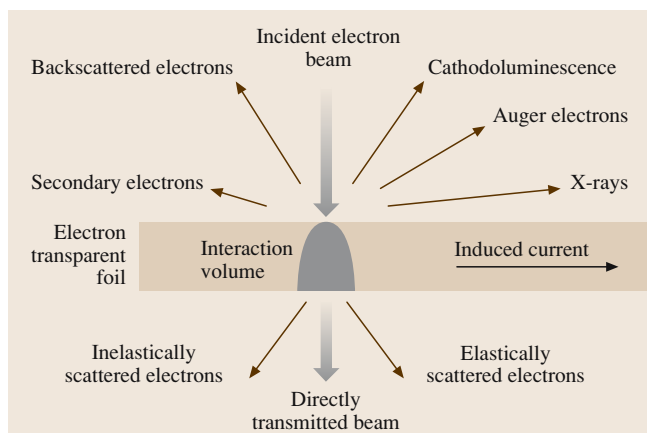


Fig. 17.4 Schematic illustrating the signals produced by the interaction of an energetic electron beam with a thin sample foil

Table 17.1 Overview of characterization techniques

Technique	Primary beam	Energy	Signals detected	Assessment	Spatial resolution	Elements detected	Detection limit
AES	Electron	0.5–10 keV	Auger electron	Surface composition	lateral \approx 200 nm (LaB ₆ source) lateral \approx 20 nm (FE source) depth \approx 2–20 nm	Li–U	\approx 0.1–1 at % (sub-monolayer) accuracy \approx 30% \approx 0.001–10 at %
RBS	Ion (He atoms)	> 1 MeV	Ion (He atoms)	Depth composition & thickness	lateral \approx 1 mm depth \approx 5–20 nm	B–U	
SEM / EDS	Electron	0.3–30 keV	Electron (SE, BSE) X-ray (characteristic)	Surface morphology & composition	\approx 1–5 nm (SE) < 1 μ m (BSE) lateral > 0.3 μ m (EDS) depth \approx 0.5–3 μ m (EDS)	B–U	\approx 0.1–1 at % accuracy \approx 20% (depends on matrix)
SIMS	Ion	1–30 keV	Ion (secondary)	Depth trace composition	lateral \approx 60 μ m (Dynamic SIMS) lateral \approx 1 μ m (Static SIMS) depth \approx 2–20 nm	H–U	\approx 10 ^{–10} – 10 ^{–5} at. %
TEM / EDS / EELS	Electron	100–400 keV	Electron (elastic, inelastic) X-ray (characteristic)	Structure and chemistry of thin sections (high resolution)	\approx 0.1–0.3 nm lateral > 2 nm (EDS) lateral \approx 1 nm (EELS) energy resolution \approx 1 eV (EELS)	up to U	\approx 0.1–1 at % accuracy \approx 20% (depends on matrix)
XPS	X-ray	1–10 keV	Photoelectron	Surface composition (chemical bonding)	lateral \approx 10 μ m–2 mm depth \approx 1–10 nm	Li–U	\approx 0.1–1 at % (sub-monolayer) accuracy \approx 30%
XRD	X-ray	1–10 keV	X-ray	Structure	lateral \approx 10 μ m depth \approx 0.1–10 μ m	Low Z may be difficult to detect	\approx 3 at % in a two-phase mixture (\approx 0.1 at % for synchrotron) accuracy \approx 10%
XRF	X-ray	30 kV / 20 mA	X-ray (fluorescent)	Composition	lateral \approx 0.1–10 mm depth \approx 10 mm	Na–U	\approx ppb - ppm, accuracy \approx 10%

lead to the discrete emission of X-ray photons characteristic of the atomic number of the element concerned. This is referred to as the K, L, M, N series, and Moseley's law states that the square root of the frequency of the characteristic X-rays of this series, for certain elements, is linearly related to the atomic number. Discrimination of these characteristic X-rays, as a function of energy or wavelength, provides the basis for the EDX or the WDX techniques, respectively. Alternatively, a secondary process of Auger electron emission may again occur, particularly for low atomic number materials, whereby outer electrons are ejected with a characteristic kinetic energy. It should also be noted that characteristic X-rays may also induce the emission of lower energy X-rays within a sample, and this is the basis of X-ray fluorescence (XRF) and the origin of background scintillation during EDX analysis.

In addition, if an outer (valence) electron state is vacant, relaxation across the band-gap of a semiconductor may occur with the emission of light, and this

constitutes the basis of the CL technique. Alternatively, a current may be induced within a sample and non-radiative recombination pathways in the presence of structural defects (and a collection junction) provide the contrast mechanism for the EBIC technique that profiles the electrical activity within a crystalline semiconductor. Also, incident electrons that interact with atomic nuclei may become backscattered with energies comparable to the incident energy, and used to image a sample surface with a contrast that is dependent on the average local composition. Secondary (low-energy, < 50 eV) electrons (SEs), produced by a variety of mechanisms, may also be emitted and escape from the near-surface of a sample. SEs can be used to obtain topographic images of irregular surfaces since they are easily absorbed.

Table 17.1 lists the most commonly used characterization techniques based on X-ray, ion or electron beam interactions with a semiconductor, and broadly indicates their limits of applicability.

17.3 X-Ray Diffraction

The basic concepts behind radiation–material interactions and scattering link into the concept of diffraction, whereby the spatial distribution and intensity of scattered X-rays or electrons provide information on the arrangement of atoms in a periodic sample. The theory of wave–particle duality proposes that an electron may be considered to be a wave rather than a particle when

discussing diffraction. Electrons are scattered by electric fields within a crystal whilst X-rays are scattered by shell electrons. Nevertheless, the geometry of diffraction is very similar in both cases, being governed by Bragg's law.

The basic principles associated with diffraction (in reflection, transmission or glancing angle geometry) are generally introduced with reference to X-ray scattering and interference. X-rays are a form of energetic electromagnetic radiation of wavelength $\approx 10^{-10}$ – 10^{-11} m, of comparable size to the spacing of atoms within a solid. A crystal lattice comprises a regular array of atoms; the electron clouds around these act as point sources for spherical X-ray wavelets, through a process of absorption and re-emission, when interaction with an incident beam of X-ray occurs. Constructive interference of the scattered wavelets, termed the Huygens's principle (Fig. 17.5), occurs in preferred directions, depending on the Bravais lattice of the crystal and the X-ray wavelength. The positions of the resultant maxima in scattering intensity may be used to deduce crystal plane spacings and hence the structure of an unknown sample. Geometrical considerations show that the scattering angles corresponding to diffracted intensity maxima can be described by the Bragg equation $n\lambda = 2d \sin \theta$ (sometimes expressed as $\lambda = 2d_{hkl} \sin \theta_{hkl}$). Formally,

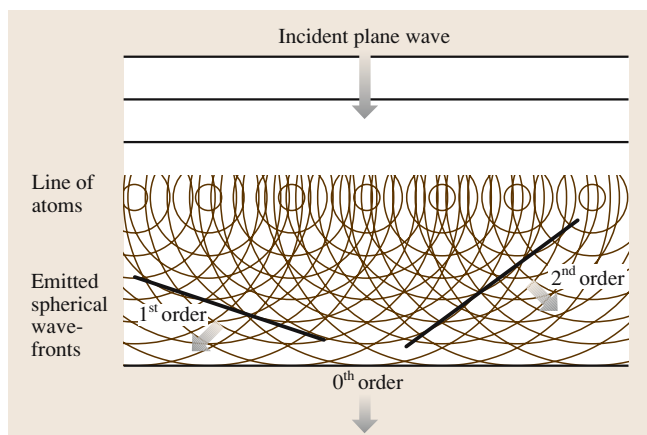


Fig. 17.5 Schematic illustrating Huygens's principle, with a reconstruction of spherically emitted X-ray wavefronts providing diffracted intensity in specific directions

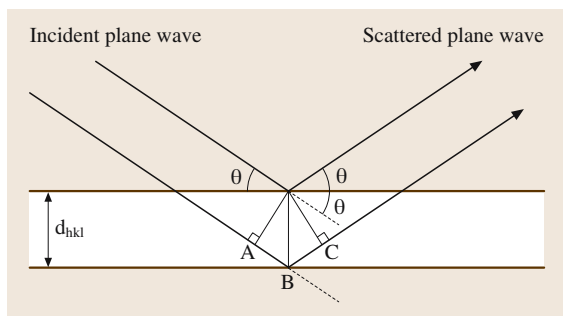


Fig. 17.6 Geometric illustration of Bragg diffraction. For constructive interference, $n\lambda = AB + BC = 2d \sin \theta$

this equation describes the minimum condition for the coherent diffraction of a monochromatic X-ray beam from a set of planes of a primitive lattice. Figure 17.6 illustrates the geometrical conditions associated with Bragg diffraction from a set of $\{hkl\}$ planes spaced d_{hkl} apart, with X-rays incident at a Bragg angle θ being diffracted through an angle 2θ . The path difference between the X-rays ‘reflected’ from successive planes must be equivalent to an integer number of wavelengths n for constructive interference to occur.

In order to interpret the information contained in the *intensity* of the diffracted beam, which is measured in practice, it is necessary to consider the amplitude of the elastically scattered waves. (The phase of the scattered beam is much more difficult to measure.) A useful concept to introduce at this stage is that of the atomic scattering factor, f , which is a measure of the amplitude of the wave scattered by an atom, which depends on the number of electrons in the atom. Formally, the atomic scattering factor is the ratio of the amplitude of the wave scattered by an atom to the amplitude scattered in the same direction by a free classical electron. (For the purpose of such discussions, an electron orbiting an atom is considered to be a free classical electron.) The scattered amplitude (and hence intensity) varies with direction, with higher angles θ having lower amplitudes.

When X-rays interact with a periodic crystal lattice, it is considered that each atom scatters with an amplitude f into an hkl reflection, while there is a summation of all of the scattering amplitudes from different atoms, with phase differences that depend on hkl and the relative positions of the atoms. This leads to the concept of the structure factor, F_{hkl} , which is the total scattering amplitude from all of the atoms in one unit cell of a lattice. Formally, the structure factor is the ratio of the amplitude scattered by a unit cell into an hkl reflection to that scattered in the same direction by

a free classical electron. The phase difference between waves scattered from two different atoms depends on the Miller indices of the reflection being considered and the fractional coordinates of the atoms within the unit cell. In general terms, for a reflection from a set of $\{hkl\}$ planes, the phase difference ϕ between the wave scattered by an atom at the origin and that scattered by an atom with fractional coordinates x, y, z is given by $\phi = 2\pi(hx + ky + lz)$. In theoretical terms, the resultant intensity of the scattered beam is denoted $I_{hkl} \propto |F_{hkl}|^2$, where $F_{hkl} = \sum f_j(\theta) \exp(i\phi_j)$, which is a summation of the individual scattered sinusoidal waves, performed over both phase and amplitude. In a practical diffraction experiment, however, the combination of photoionization and Compton scattering can act to diminish the scattered beam intensity. Consideration also needs to be given to the effects of absorption, along with the effect of multiplicity, which arises from the number of symmetrical variants of a unit cell, and the geometrical and polarization factors specific to a given experimental arrangement.

The crystallographic structure of an unknown material can, nevertheless, be analyzed via the diffraction of X-rays of known wavelength. For example, for a crystal system with orthogonal axes, the general formula which relates plane spacing, d_{hkl} , to the plane index $\{hkl\}$ and the lattice parameters a, b, c is $1/(d_{hkl})^2 = h^2/a^2 + k^2/b^2 + l^2/c^2$. For a cubic system, this simplifies to $d_{hkl} = a/\sqrt{(h^2 + k^2 + l^2)} = a/\sqrt{N}$. Thus, $\lambda = 2a \sin \theta / \sqrt{N}$, from which $N = (4a^2/\lambda^2) \sin^2 \theta$ and hence $N \propto \sin^2 \theta$. Accordingly, the 2θ angles of scattering arising from the process of X-ray diffraction may be used to identify values for N , against which hkl indices can be assigned and the lattice identified. Information is contained within the intensities and widths of the diffracted X-ray peaks, in addition to their positions. The process of formal identification can be automated by electronically referencing crystallographic databases containing the positions and relative magnitudes of the strongest diffraction peaks from known compounds. Systematic or partial absences may also occur depending on the lattice symmetry, and the intensities of some peaks may be weak due to the motif of the crystal lattice – the number of atoms sited on each lattice point. For example, sphalerite or wurtzite lattices may show weak reflections that are absent for face-centered cubic or hexagonal close-packed structures, respectively.

Partial differentiation of the Bragg equation gives $2(\sin \theta)\delta d + 2d(\cos \theta)\delta \theta = 0$, from which $\delta d/\delta \theta = -d \cot \theta$. Thus, for a fixed error in θ , the error in d_{hkl} will minimize as $\cot \theta$ tends to zero (as θ tends

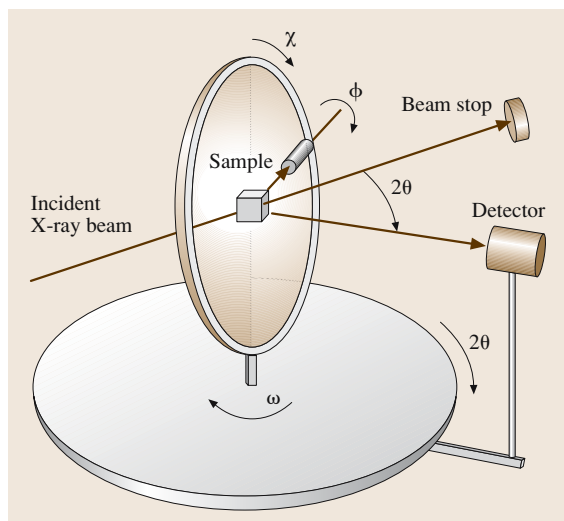


Fig. 17.7 X-ray four-circle diffractometer

to 90°). Diffraction techniques are therefore more accurate when measurements are made at high 2θ angles. Similarly, it can be shown that the sensitivities and resolutions of diffraction measurements improve at high angles. The sensitivity of the X-ray diffraction technique, for example, is sufficiently high to enable subtle stress measurements to be made, with elastic changes in plane spacing leading to small shifts in diffracted peak positions. Accordingly, diffraction techniques may be used to investigate temperature-dependent order–disorder transitions within alloys and preferred orientation effects. The high accuracy of lattice parameter measurement similarly enables good compositional analysis of an alloy, for example by assuming Vegard’s law, that the alloy exhibits a linear dependence of lattice parameter on composition between the extremes of composition. Brief reference should be made here to the related technique of neutron diffraction, which is able to sample over a much larger range of d_{hkl} spacings, which makes it particularly useful for the *ab initio* determination of very complex structures. The very rigorous approach of Rietveld refinement can also be used to model a complete diffraction profile, deducing the structure from first principles using atom positions as fitted parameters.

The principle behind the generation of X-rays from characteristic transitions, used when performing chemical microanalyses of unknown samples by EDX or WDX, also provides the route to defining an X-ray source from a known target sample for the purpose of XRD. In practical terms, the X-ray tube of a diffrac-

tometer is an evacuated vessel in which electrons from a hot filament are focused onto a cooled metal target, such as Cu or Ni. The X-ray spectrum generated consists of X-rays characteristic of the target material, along with a background emission of X-rays of continuous wavelengths, termed Bremsstrahlung (braking radiation), which are caused by the acceleration of electrons in the vicinity of nuclei. The X-rays emerging from the tube, through a window made from a material of low atomic number, can be filtered and collimated to define a beam of specific wavelength.

The precise geometry used in a diffraction experiment will depend on the form of the sample and the information content required. For example, crystalline samples such as semiconductor heterostructures can be rotated within a cylindrical geometrical framework that enables successive sets of crystal planes to be brought into play for detection. Figure 17.7 illustrates the basic geometrical arrangements for a four-circle diffractometer that allows diffraction spectra to be acquired. Computer-controlled rotation of the sample around ω , χ and ϕ axes, as the detector is rotated about the 2θ axis, enables the positions and intensities of hkl reflections to be recorded. By way of example, Fig. 17.8 shows a $2\theta/\omega$ plot recorded from a heteroepitaxial GaN/GaAs(001) sample, with reflections attributable to both epilayer and substrate, whilst the full width at half maximum (FWHM) values for the diffraction peaks provide a measure of the mosaic spread of the subgrains within the sample.

Since amorphous materials do not exhibit long-range order, their diffraction profiles show diffuse intensities rather than well-defined maxima. Partially crystalline materials may show broad diffraction peaks, from which

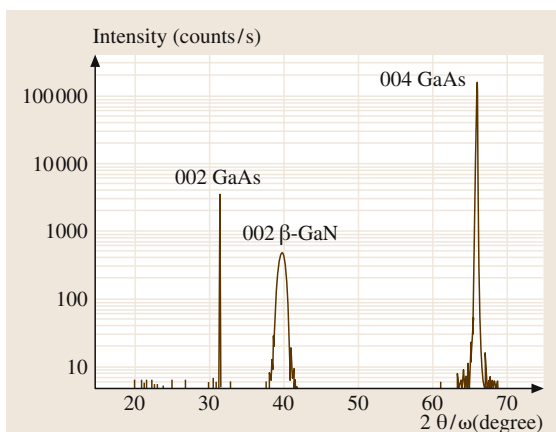


Fig. 17.8 Example of a $2\theta/\omega$ plot for cubic GaN/GaAs(001)

it is possible to approximate the crystal size from the λ -peak-width relationship. Other X-ray diffraction techniques include Laue back reflection, that can be used conveniently to orient bulk single crystals, for example for sectioning prior to use as substrates for heteroepitaxial growth. Alternatively, the Debye–Scherrer method can be used for powder samples, since a significant number of crystal grains will always be in an orientation that satisfies the Bragg equation for each set of $\{hkl\}$ planes. In this scattering arrangement, the diffracted rays form cones coaxial with the incident X-ray beam, with each cone of diffracted rays corresponding to a Bragg reflection from a specific set of lattice planes in the sample. A cylindrical strip of photographic film can be used to detect the diffracted intensity.

To reiterate, it is the combination of Bragg's law and the structure factor equation that enables the directions and intensities of beams scattered from a crystal to be predicted. In this context, it is instructive to briefly compare XRD with electron diffraction. Electrons are scattered by the periodic potential – the electric field – within a crystal lattice, whilst X-rays are scattered by shell electrons. Since X-rays and electrons exhibit comparable and comparatively small wavelengths, respectively, on the scale of the plane spacings of a crystal lattice, this equates to large and small angles of scatter-

ing, respectively, for the diffracted beams. Accordingly, in principle XRD techniques offer greater accuracy than electron diffraction for the measurement of lattice parameters. It should also be noted that XRD is essentially a kinematic process based on single scattering events, whilst electron diffraction is potentially more complex due to the possibility of dynamic (or plural) scattering processes which can affect the generated intensities. Also, electrons are more strongly absorbed than X-rays, so there is need for very thin sample foils, typically $< 1\ \mu\text{m}$, for the purposes of transmission electron diffraction (TED) experiments. However, electrons are more easily scattered by a crystal lattice than X-rays, albeit through small angles, so an electron-transparent sample foil is capable of producing intense diffracted beams. X-rays require a much greater interaction volume to achieve a considerable diffraction intensity. The effectiveness of the technique of electron diffraction becomes most apparent when combined with TEM-based chemical microanalysis imaging techniques. This enables features such as small grains and embedded phases, or linear or planar defect structures such as dislocations and domain boundaries, to be investigated in detail.

Before describing some variants of the electron diffraction technique, a few concepts related to imaging and modes of operation of the TEM need to be introduced.

17.4 Optics, Imaging and Electron Diffraction

The aim of a microscope-based system is to image an object at high magnification, with optimum resolution and without distortion. The concepts of magnification and resolution associated with imaging in electron microscopy are usually introduced via light ray diagrams for optical microscopy. The constraints on achieving optimum resolution in TEM are generally considered to be lens aberration and astigmatism. The concepts of depth of field and depth of focus must also be considered.

If we consider the objective lens shown in Fig. 17.9, a single lens is characterized by a focal length f and a magnification M . The expression $1/f = 1/u + 1/v$ relates the focal length to the object distance u and the image distance v for a thin convex lens. The magnification of this lens is then given by $M = v/u = f/(u - f) = (v - f)/f$, from which it is apparent that $u - f$ must be small and positive for a large magnification to be obtained. In practice, a series of lenses are used to achieve a high magnification overall

whilst minimizing distortion effects. For the combined projection microscope system shown in Fig. 17.9, magnification scales as $M' = (v - f)(v' - f')/ff'$.

Resolution is defined as being the smallest separation of two points on an object that can be reproduced distinctly within an image. The resolution of an optical lens system is diffraction-limited since light must pass through a series of apertures, and so a point source is imaged as a set of Airy rings. Formally, the minimum resolvable separation of two point sources, imaged as two overlapping sets of Airy rings, is given by the Rayleigh criterion, whereby the center of one set of Airy rings overlaps the first minimum of the second set of Airy rings. The defining equation for resolution is given by $r = 0.61\lambda/n \sin \alpha$, where λ is the wavelength of the imaging radiation, n is the refractive index of the lens, and α is the semiangle subtended at the lens. The combined term $n \sin \alpha$ is the 'numerical aperture' of the lens. Thus, resolution can be improved by decreasing λ

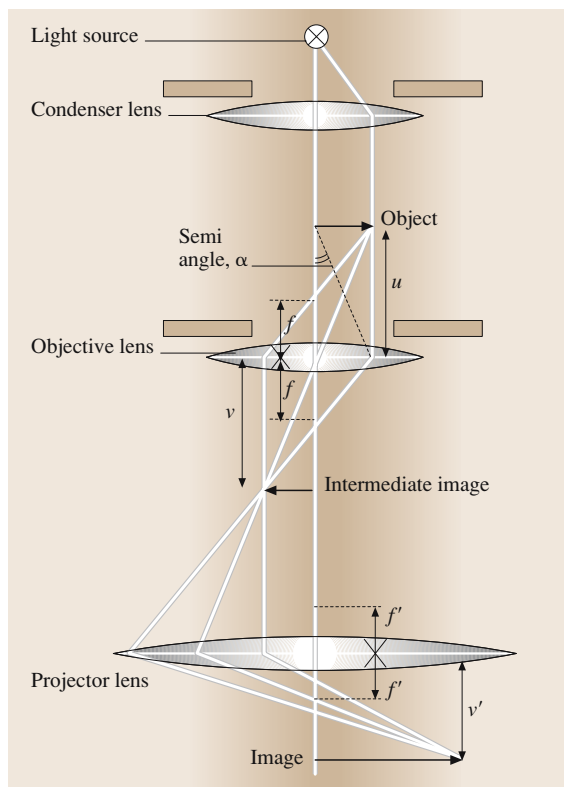


Fig. 17.9 Schematic ray diagram for a transmission projection optical microscope

or by increasing n or α . It is noted that magnification beyond the resolution limit of the system is possible, but there is no benefit to achieving this, and so it is termed ‘empty magnification.’

The performance of a **TEM** can be described in analogous terms, comprising an electron gun and a series of electromagnetic lenses for sample illumination, image formation and magnification: the condenser, objective and projector systems, respectively (Fig. 17.10). The electron source can be either a hot filament (e.g. W or LaB₆) for thermionic emission, or a hot or cold cathode (e.g. ZrO₂ coated W) for field emission. The gun and lenses are traditionally assembled in a column with a viewing screen and camera at the bottom to record images or diffraction patterns from an electron-transparent sample placed within the objective pole piece. A range of differently sized apertures are located in the condenser system to help collimate the electron probe, in the projector system to select a region of sample from which a diffraction pattern may be formed, and in the objective lens just below the sample to se-

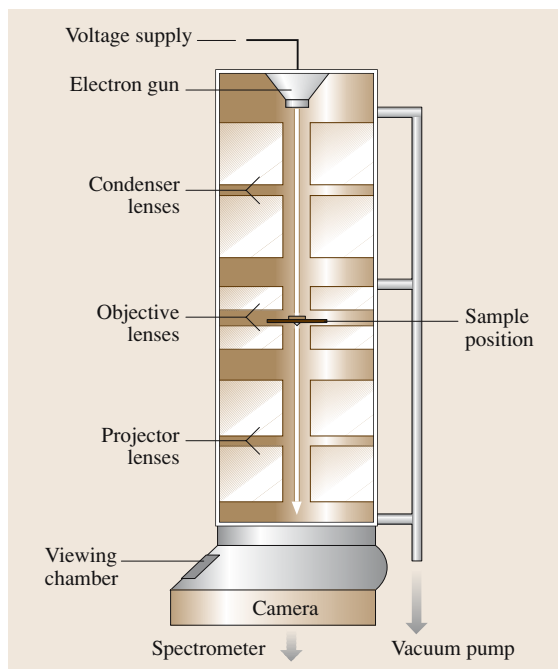


Fig. 17.10 Simplified lens configuration of a **TEM**

lect the transmitted beam(s) used to form an image. The path of travel of the electron beam through the entire electron-optic column must be under conditions of high vacuum, considering the ease of absorption of electrons in air. The electrons are accelerated by high voltage, typically 100 or 200 kV, although there are **TEMs** that operate at MV conditions, depending on the intended application. An accelerated, high-energy electron acquires significant kinetic energy and momentum. It can also be represented by a wavelength (corrected to take into account relativity) that can be approximated in nanometers by $\lambda = [1.5/(V + 10^{-6} V^2)]^{0.5}$. By way of example, an electron beam within a conventional **TEM** operating at 200 kV has a relativistically corrected wavelength of 0.0025 nm, as compared with the range of ≈ 400 –700 nm for visible light. Since resolution is wavelength-limited, the technique of **TEM** is, in principle, able to provide a vast improvement in resolution over conventional optical systems. The equation for resolution becomes approximated by $r \approx 0.61\lambda/\alpha$ since $n = 1$ for electromagnetic lenses and $\sin \alpha \approx \alpha$ in view of the small angles associated with electron diffraction. A value of $\alpha \approx 150$ mrad would suggest a resolution approaching ≈ 0.01 nm. However, this level of resolution for **TEM** is generally not achievable in practice due to the effect of lens aberration, which de-

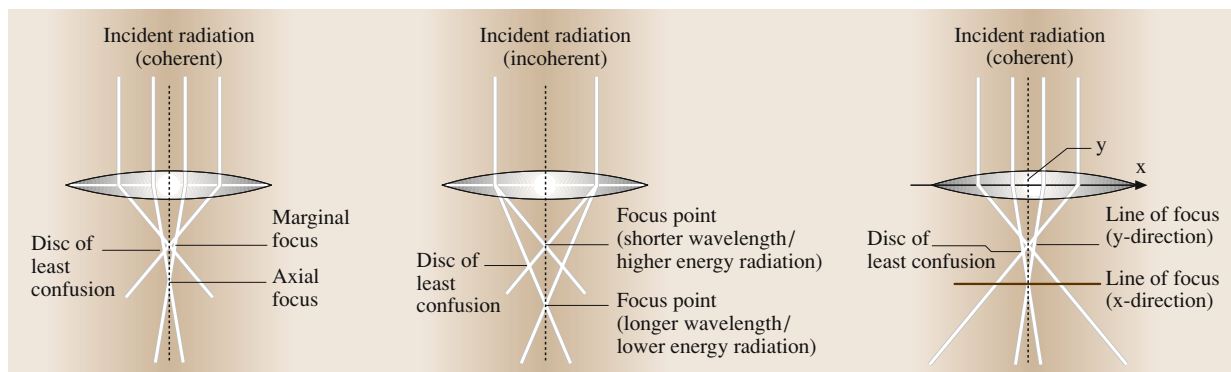


Fig. 17.11a–c Schematic illustration of (a) spherical aberration; (b) chromatic aberration and (c) astigmatism

grades the resolution of an image to a ‘disc of least confusion’ above the theoretical wavelength-limited resolution of the electron-optic system. Resolution values of $\approx 0.2 \text{ nm} - 0.3 \text{ nm}$ are associated with the best conventional instruments. (When charting the historical development of TEM, improvements in resolution have depended on the construction of microscopes operating at higher voltages, in order to capitalize on the benefits from the reduced electron wavelength. However, it should be noted that recent improvements in techniques of compensating for spherical aberration have now enabled a resolution of $\approx 0.1 \text{ nm}$ to be achieved for intermediate-voltage, field emission gun instruments.)

The variation in focal length of a lens as a function of the distance of a beam from the center of the lens results in rays traveling further from the optic axis being brought to focus closer to the lens (Fig. 17.11a). This spread in path lengths of rays traveling from an object to the image plane is termed spherical aberration. In this case, the radius of the disc of least confusion is given by $r_s = C_s \alpha^3$, when referred back to the object, where C_s is termed the spherical aberration coefficient and α is again the semiangle, in radians, subtended at the lens. Spherical aberration can be limited by reducing α ; in other words by ‘stopping down’ the lens by using smaller apertures (in this case, the term α can be represented by a term β corresponding to the ‘aperture collection angle’). However, this conflicts with the large value of α needed to optimize resolution, so a balance is required for optimum resolution, which is achieved when $\alpha_{\text{opt}} \sim (\lambda/C_s)^{1/4}$, corresponding to an effective resolution of $r_{\text{opt}} \sim \lambda^{3/4} C_s^{1/4}$. Lenses also exhibit different focal lengths for electrons with different wavelengths; such a spread of wavelengths can arise from slight fluctuations in the accelerating voltage or from

inelastic scattering (energy loss) processes within the specimen, and this effect is termed chromatic aberration (Fig. 17.11b). The radius of the disc of least confusion in this instance is given by $r_c = C_c \alpha (\Delta E/E_0)$, where C_c is the chromatic aberration coefficient, E_0 the accelerating voltage and ΔE the spread in electron energy.

An additional aberration, termed astigmatism, arises from the asymmetry of a lens about the optic axis. The different focal lengths of a lens for different orientations leads to a loss of sharpness of the image at focus.

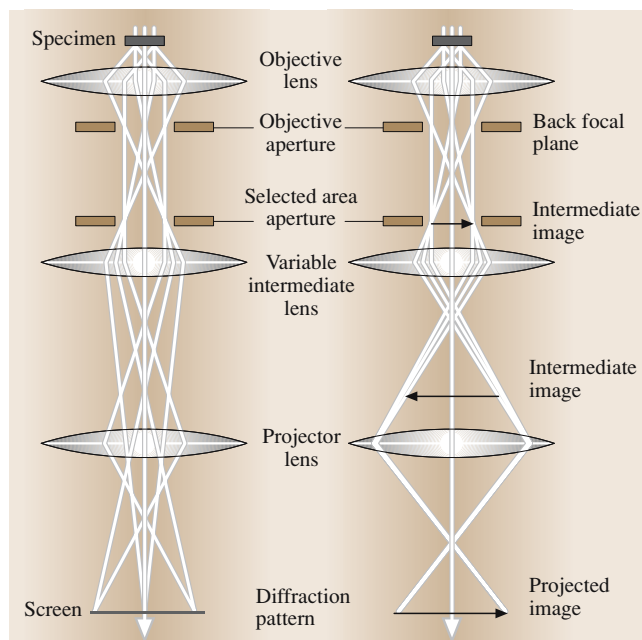


Fig. 17.12a,b Schematic ray diagrams showing the projection of a (a) diffraction pattern and (b) an image onto a TEM viewing screen

One measure of the astigmatism is the distance between the two foci formed at right angles along the optic axis (Fig. 17.11c), in contrast to a single point of focus. Astigmatism may be compensated for by using electromagnetic stigmators that generate a compensating field to bring the rays back into common focus. Distortion of an image can also occur due to slight variations in magnification with radial distance from the optic axis, leading to so-called ‘pin-cushion’ or ‘barrel’ distortion. This effect can become particularly noticeable at very low magnification.

The term ‘depth of field’ is defined as the distance along the optic axis that an object can be moved without noticeably reducing the resolution. This effect is again dependent on the radius of the disc of least confusion that can be tolerated and α . The depth of field approximates to λ/α^2 , and a typical value of a few tens of nm for electron microscopy means that every point within the thickness of a typical electron-transparent foil can be imaged at focus. Conversely, the ‘depth of focus’ corresponds to the maximum permissible spacing between the imaging screen and the photographic plate or CCD used to record an image. The depth of focus approximates to $\lambda M^2/\alpha^2$, and since this works out at many meters, the viewing screen and recording system of an electron microscope need not coincide.

Following on from these general considerations affecting the process of imaging, we now move on to briefly describe the conventional modes of operation of a TEM. The accelerating stack and condenser system of the microscope defines the high-energy electron probe incident at a thin sample, within which many complex interactions occur and various signals are produced, as detailed earlier. A diffraction pattern is initially formed in the back focal plane of the objective lens and the recombination of diffracted beams allows the reconstruction of an inverted image in the first image plane. Changing the strengths of the intermediate lenses allows the back focal plane (corresponding to a projection of reciprocal space) or image plane (corresponding to a projection of real space) to be observed on the viewing screen (Fig. 17.12a,b). Placing an aperture with a selected area located in an intermediate image plane around an area of interest within a projected image enables us to ensure that only beams from that particular area of the sample contribute to the diffraction pattern viewed on the screen.

It is sometimes convenient to think of the intersection of an Ewald sphere with a reciprocal lattice when describing the construction and projection of a diffraction pattern. A reciprocal lattice is constructed from

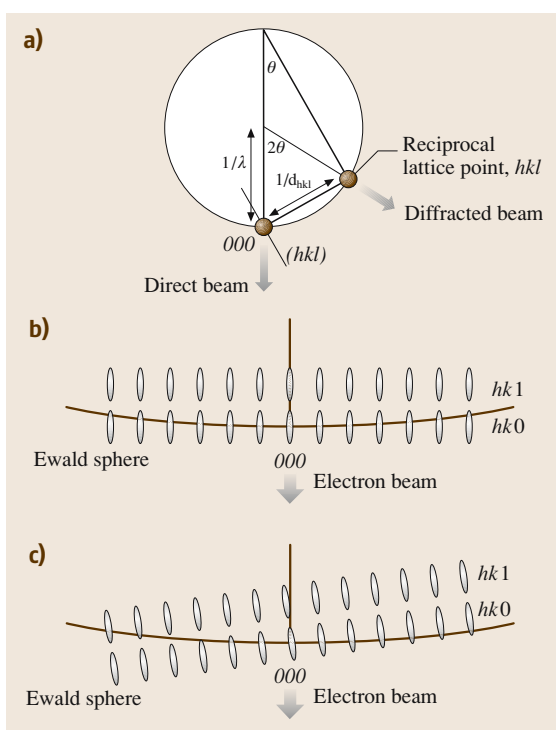


Fig. 17.13a–c Ewald sphere construction for (a) X-ray diffraction (large angle scattering); (b) electron diffraction (zone axis projection) and (c) electron diffraction (tilted projection)

a crystal lattice such that any vector from the origin to a diffracted spot is normal to a particular plane in the real lattice, with a reciprocal length given by the plane spacing. Thus, a three-dimensional crystal lattice can be transformed into a three-dimensional reciprocal lattice. The Ewald sphere can be thought of as a geometrical construction of radius $1/\lambda$ intersecting the reciprocal lattice. Figure 17.13a illustrates the construction of an Ewald sphere for the case of large-angle scattering (X-ray diffraction). The direction of the incident beam at the sample corresponds to the direction of the radius of the Ewald sphere, and the point of emergence of this vector at the sphere surface coincident with a reciprocal lattice point is taken as the origin of the spectrum or diffraction pattern. The same construction applies for a beam of electrons, but the sphere surface has a very shallow curvature relative to the reciprocal lattice spacing, due to the very small value of λ relative to d_{hkl} (Fig. 17.13b). Also, diffraction from a thin crystal is associated with a lengthening of the reciprocal lattice spots into rods in a direction parallel to the electron beam.

Consequently, if the electron beam is incident along a low index zone axis, the Ewald sphere approximates to a plane and intersects a layer of reciprocal lattice points, and a two-dimensional array of diffraction spots is projected. In very general terms, it is considered that a set of $\{hkl\}$ planes is at the Bragg condition when the reciprocal lattice point corresponding to hkl falls on the surface of the Ewald sphere. For the case of large angles of diffraction, an outer ring of diffraction spots may be observed, termed a high-order Laue zone (HOLZ), since the Ewald sphere has sufficient curvature to intersect with a neighboring layer of points within the reciprocal lattice (Fig. 17.13c).

We should also mention the Kikuchi lines that arise in diffraction patterns due to the elastic scattering of incoherently scattered electrons. The intensities of these Kikuchi lines increase with increasing thickness of the sample foil, and the line spacings are the same as the spacings of the diffraction spots from the associated crystal planes. Kikuchi lines move as the sample is tilted and hence can be used to establish very precise crystal orientations for the purpose of image contrast or convergent beam electron diffraction (CBED) experiments.

Referring back to Fig. 17.12, a diffracted beam may be selected using an aperture inserted into the back focal plane of the objective lens and used to form an image. If the undeviated, transmitted beam is used, then a bright field image is formed where the areas that diffract strongly appear dark. Image contrast also arises from a mass-thickness effect, whereby thicker or high-density regions of material scatter more strongly and hence appear dark. Alternatively, centered dark field images can be created by aligning a diffracted beam, termed a g_{hkl} reflection, down the optic axis of the microscope (in contrast to moving the aperture over the diffraction spot, which would increase aberrations and degrade resolution). Areas where the $\{hkl\}$ planes diffract strongly appear bright in such cases. Weak beam images can also be produced under dark field imaging conditions, with the sample tilted slightly away from a strong Bragg condition. The trade-off, in this instance, is reduced image contrast for improved resolution, which allows fine detailed features, such as partial dislocations bounding dissociated dislocations, to be delineated.

By way of example, Fig. 17.14 shows the different contrasts obtained from a plan-view $\text{Si}/\text{Si}_{0.96}\text{Ge}_{0.04}/\text{Si}(001)$ sample foil imaged under bright-field and weak-beam diffraction conditions, respectively. The sample foil contains orthogonal arrays

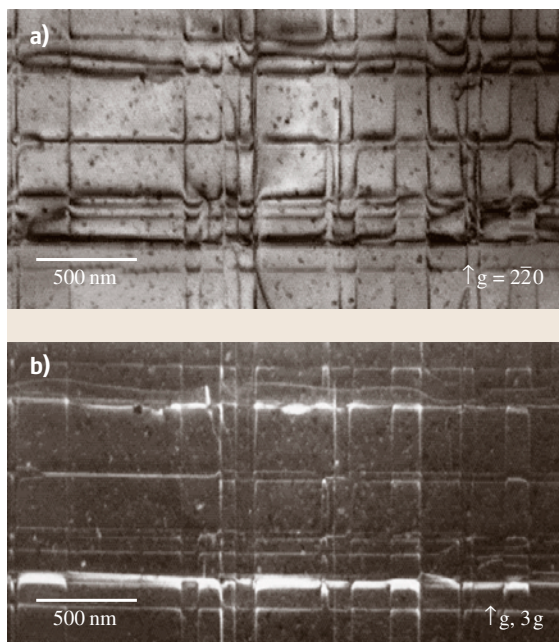


Fig. 17.14 (a) Bright-field and (b) weak-beam plan-view TEM images of $\text{Si}/\text{SiGe}/\text{Si}(001)$

of interfacial misfit dislocations that are generated as a consequence of the relaxation of the strain arising from the different lattice parameters of Si and $\text{Si}_{0.96}\text{Ge}_{0.04}$ [17.16]. The strain fields around the dislocations lead to a local deviation away from the exact Bragg condition, which can be used to delineate the (approximate) positions of the dislocation cores. The weak beam technique allows the complex dislocation tangle to be resolved more clearly. Imaging such sample foils using different diffraction vectors produces different contrasts that allow the nature of the dislocations to be precisely determined. Hence, the dislocation reaction mechanisms responsible for microstructure development can, in principle, be identified.

Some examples are now presented that illustrate how electron diffraction can contribute to the structural characterization of semiconductors. The use of conventional TED combined with image contrast analysis to assess the defect microstructure of a semiconductor heterostructure is initially considered. The application of the CBED technique, sometimes termed microdiffraction, in order to determine the polarities of noncentrosymmetric crystals is then described. We also focus on the RHEED technique, which may be used for rapidly assessing the near-surface microstructures of semiconductor thin films.

17.4.1 Electron Diffraction and Image Contrast Analysis

Figure 17.15 shows a TED pattern corresponding to a highly symmetric, low-index, zone axis projection, acquired from an electron-transparent foil of epitaxial GaN grown on the {0001} basal plane of sapphire, viewed in cross-section under spread beam conditions, in other words near-parallel illumination.

It is generally instructive to view such diffraction patterns alongside other images of the sample. Figure 17.16 compares high-resolution electron microscopy (HREM) and conventional TEM (CTEM) images of heteroepitaxial GaN–sapphire that were acquired for this projection. The HREM mode of imaging, otherwise termed ‘phase contrast’ or ‘lattice imaging’, makes use of several diffracted beams selected by a large objective aperture; the resulting interference patterns can be used to elucidate the locations of atomic columns (Fig. 17.16a). Careful simulation is required to precisely assign the atomic positions within such images, since the contrast is strongly dependent on defocus,

foil thickness and sample orientation. However, the structural integrity of the interface between GaN and sapphire can be appraised, along with the presence of nanometer-scale, three-dimensional growth islands formed during the initial stages of epitaxy, prior to layer coverage. Conversely, the conventional ‘many-beam, bright-field’ image of heteroepitaxial GaN–sapphire (Fig. 17.16b) was created using a small objective aperture placed over the directly transmitted beam, and recorded at lower magnification. The image contrast is again complicated because there are many excited diffracted beams operating for this low-index crystal orientation and strain fields associated with the large number of threading defects ($\approx 10^{10} \text{ cm}^{-2}$) within this sample. Despite this level of defect content, the diffraction pattern of Fig. 17.15 still reflects the high level of crystallographic perfection of the matrix. Also, rotation of this sample foil by 30° about the growth axis enabled a $\langle 1\bar{1}00 \rangle_{\text{nitride}} || \langle 11\bar{2}0 \rangle_{\text{sapphire}}$ projection to be established, consistent with the hexagonal symmetry of this GaN–sapphire system. Indeed, establishing different diffraction patterns for different sample projections, for known angles of sample tilt, enable the phase and structural relationship of an epilayer and substrate to be readily established.

There is a need to characterize the precise nature of the fine-scale defect content within such samples. Dislocations, for example, generally act as nonradiative recombination centers, which can deleteriously affect the charge transport properties of a semiconductor [17.17]. Gaining an improved understanding of how dislocations are created and how they interact enables us to identify mechanisms that could be used to control their development, thereby improving growth of semiconductor material and the resultant properties of devices made from it.

Dislocations are one-dimensional defects that can be pure edge, pure screw or ‘mixed edge-screw’ in character. In certain circumstances, dislocations can dissociate

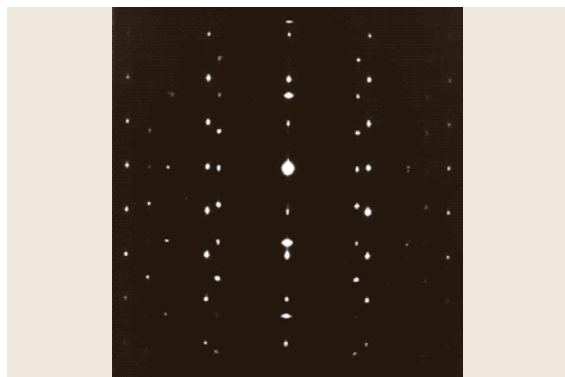


Fig. 17.15 TED pattern acquired from epitaxial GaN/sapphire{0001} viewed in cross-section, corresponding to the $\langle 11\bar{2}0 \rangle_{\text{nitride}} || \langle 1\bar{1}00 \rangle_{\text{sapphire}}$ axis projection

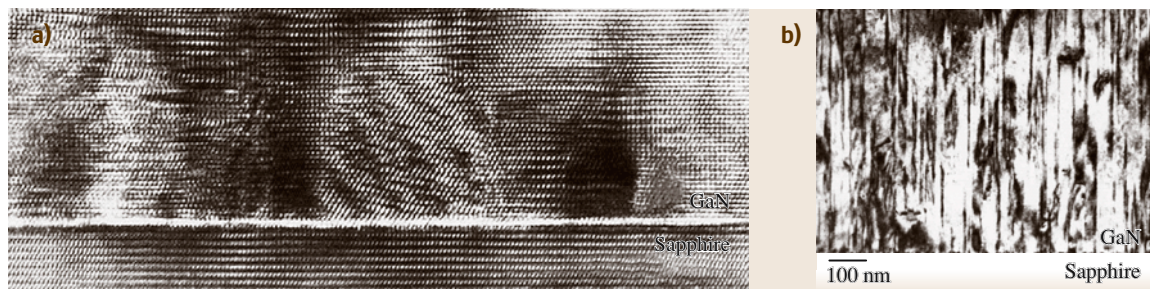


Fig. 17.16 (a) HREM and (b) many-beam CTEM images of GaN/sapphire{0001}

to form a pair of partial dislocations bounding a stacking fault ribbon, the separation of which depends on the material stacking fault energy. Dislocations can be described with reference to their line direction \mathbf{u} and Burgers vector \mathbf{b} , representing the closure failure of a loop of equal atomic spacing around the defect core. Dislocations can move through a crystal lattice by glide or climb mechanisms; for example, when under deformation or as part of a strain relaxation mechanism during heteroepitaxial growth. Also, dislocations can interact to either self-annihilate or multiply to generate more dislocations depending on their respective type, habit plane and the slip system of the matrix.

The technique of diffraction contrast analysis allows the nature of dislocations to be ascertained. The approach used is to tilt the sample away from the highly symmetrical, low-index zone axis orientation in order to establish diffraction contrast images under selected “two-beam” conditions, with one strong diffraction spot excited (\mathbf{g}) in addition to the central transmitted beam, corresponding to one set of crystal planes at the Bragg condition. When $\mathbf{g} \cdot \mathbf{b} = 0$, the displacement associated with a dislocation does not affect the diffracting planes used to form the image and so the defect appears invisible. In practice, two examples of this invisibility condition are generally required to determine the precise displacement associated with a given defect. The dislocation will appear visible when $\mathbf{g} \cdot \mathbf{b} = 1$ and might show a more complex double image for the case of $\mathbf{g} \cdot \mathbf{b} \geq 2$. Since a screw dislocation is characterized by a Burgers vector \mathbf{b} parallel to the line direction \mathbf{u} , the defect is invisible if a diffraction vector perpendicular to the line direction is chosen. In the case of an edge or mixed dislocation where \mathbf{b} is not parallel to \mathbf{u} , there is the stricter requirement for both $\mathbf{g} \cdot \mathbf{b} = 0$ and $\mathbf{g} \cdot (\mathbf{b} \times \mathbf{u}) = 0$ for true invisibility, otherwise residual contrast might be present that acts to confuse the image interpretation. We should also mention the deviation parameter s when establishing a diffracting condition, which represents the distance in reciprocal space from the exact Bragg condition, since this is associated with imaging artefacts such as extinction contours and thickness fringes that may act to further complicate an image.

Bright-field, dark-field or weak-beam diffraction contrast imaging techniques can be used for defect analysis, depending on the resolution required. By way of example, epitaxial GaN-($\overline{1}\overline{1}\overline{1}$)B GaAs grown by molecular-beam epitaxy (MBE) at 700 °C exhibits a mosaic cell structure with subgrain boundaries delineated by predominantly mixed-type threading dislocations (typically $> 10^{11} \text{ cm}^{-2}$) [17.18]. The

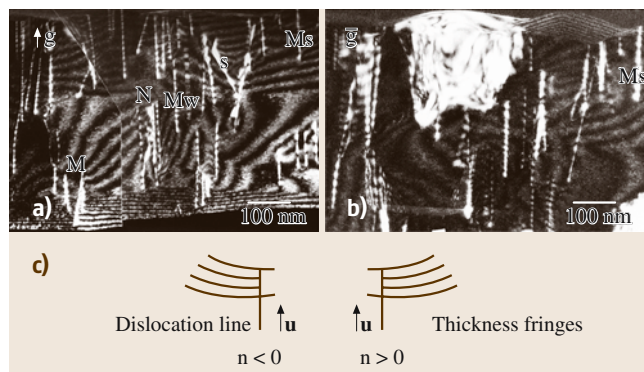


Fig. 17.17a–c Weak-beam images of the same regions of a GaN epilayer viewed in cross-section using (a) $\mathbf{g} = 000\overline{2}$ and (b) $\mathbf{g} = 1\overline{1}00$. (c) Schematic diagram summarizing Ishida’s rule for determining \mathbf{b} for $n = -2$ and $n = 2$

epitaxial relationship is given by $[000\overline{1}]_{\text{GaN}} \parallel [\overline{1}\overline{1}\overline{1}]_{\text{GaAs}}$, $[1\overline{1}00]_{\text{GaN}} \parallel [2\overline{1}\overline{1}]_{\text{GaAs}}$ and $[11\overline{2}0]_{\text{GaN}} \parallel [01\overline{1}]_{\text{GaAs}}$, so the mismatch between GaN and GaAs is 38.2% between $\{1\overline{1}00\}_{\text{GaN}}$ and $\{2\overline{2}0\}_{\text{GaAs}}$, with the epilayer in tensile strain. Figures 17.17a,b are weak-beam images of the epilayer viewed in cross-section near the $[11\overline{2}0]$ zone axis, using $\mathbf{g} = 000\overline{2}$ and $\mathbf{g} = 1\overline{1}00$, respectively. Most of the dislocations have a line direction of $\langle 0001 \rangle$. Examples of perfect edge, $\mathbf{b} = 1/3\langle 11\overline{2}0 \rangle$, and screw-type dislocations, $\mathbf{b} = \langle 0001 \rangle$, are apparent, but mostly dislocations (typically $\approx 70\%$) can be seen in both images, hence they are mixed-type, with Burger’s vector components of \mathbf{a} and \mathbf{c} , in other words $1/3\langle 11\overline{2}3 \rangle$.

It is possible to move this analysis forwards by applying Ishida’s rule, which allows the magnitude and sense of a Burgers vector to be determined if there are terminating thickness fringes at the exit of a dislocation from a wedge-shaped sample foil [17.19]. If there are n fringes terminating at one end of a dislocation, then $\mathbf{g} \cdot \mathbf{b} = n$ and the sign of n ($n > 0$ or $n < 0$) is defined according to Fig. 17.16c. For example, the screw dislocation S (Fig. 17.17a) imaged using $\mathbf{g} = 000\overline{2}$ has two thickness fringes terminating on the right-hand side of the dislocation, with the dislocation line pointing towards the $[000\overline{1}]$ growth direction. Therefore $\mathbf{g} \cdot \mathbf{b} = -2L = 2$, so \mathbf{b} is $[000\overline{1}]$ for this screw-type dislocation. For dislocations of mixed type, Figs. 17.17a,b show that some have strong contrast in both images, while others have strong contrast when imaged using $\mathbf{g} = 000\overline{2}$ and weak contrast when $\mathbf{g} = 1\overline{1}00$. For the former, for example the dislocation pointed out as M_s in Fig. 17.17a, thickness fringes terminate at the left-hand side of the dislocation with reference to the dislocation line pointing downwards

along $[0001]$. Hence, $\mathbf{g} \cdot \mathbf{b} = -2L = -2$ for $\mathbf{g} = 000\bar{2}$ and $\mathbf{g} \cdot \mathbf{b} = -1$ for $\mathbf{g} = 1\bar{1}00$, so its Burgers vector is either $1/3[\bar{2}113]$ or $1/3[\bar{1}2\bar{1}3]$. For the cases of strong contrast by $\mathbf{g} = 000\bar{2}$ and weak contrast by $\mathbf{g} = 1\bar{1}00$, such as the dislocation labeled M_w in Fig. 17.17b, such dislocations have \mathbf{b} of $\pm 1/3[11\bar{2}3]$ or $\pm 1/3[11\bar{2}\bar{3}]$, and the weak contrast is due to $\mathbf{g} \cdot (\mathbf{b} \times \mathbf{u}) \neq 0$. Because the thickness fringes connected with this defect type are not clear in the images, a more precise value for \mathbf{b} cannot be determined. It is also noted that some dislocations have opposite values of \mathbf{b} to others, such as those marked as M and N in Fig. 17.17a. These two opposite types of dislocations delineate a subgrain. The region showing bright contrast when imaged with $\mathbf{g} = 1\bar{1}00$ in Fig. 17.17b is also a misoriented subgrain, tilted close to the Bragg condition.

17.4.2 Microdiffraction and Polarity

If the electron beam converges to form a focused probe at the sample, then there will be a range of beam directions within the incident probe and within the transmitted and diffracted beams. This leads to the formation of diffraction patterns comprising discs rather than spots. The fine detail within such microdiffraction or CBED patterns contains space and point group information. Such patterns also provide a sensitive measure of lattice pa-

rameters and strain within a sample, and can also be used to assess defect type.

Two examples showing how focused probe diffraction patterns may be used to determine the absolute polarities of sphalerite and wurtzite noncentrosymmetric crystals, in situ in the TEM, are now illustrated. It is often important to know the polar orientation of a heterostructure since this can strongly influence the mode and rate of growth, the incorporation of dopants or impurities, and the development of extended defects, and hence the functional performance of the resulting device structure. The opposite polar faces of noncentrosymmetric crystals [17.20] may be distinguished in practice using appropriate chemical etchants. However, such reagents are discriminatory in their action and need to be correlated with some other experimental technique for the purpose of absolute polarity determination. For most cases of diffraction, there is no difference between the intensities of beams scattered by the hkl and $\bar{h}\bar{k}\bar{l}$ reflection planes. This indeterminacy is known as Friedel's law. However, the technique of microdiffraction coupled with a breakdown in Friedel's law allows the absolute polarities of sphalerite crystals to be determined when the anion and cation sizes are very similar (as with GaAs, ZnSe and CdTe) [17.21]. The diffraction condition shown in Fig. 17.18a corresponds to a projection tilted $\approx 10^\circ$ off a GaAs $[110]$ zone axis, arrived at fol-

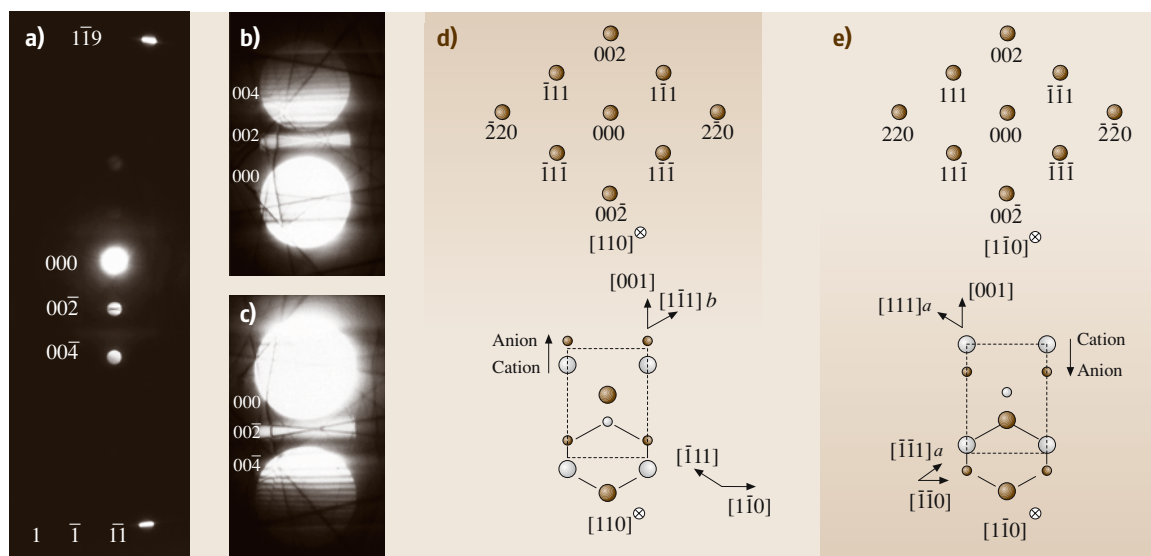


Fig. 17.18 (a)–(c) Microdiffraction patterns from $[110]$ GaAs. (a) Destructive interference through an $000\bar{2}$ diffraction disc due to the interaction of doubly diffracted $1\bar{1}9$ and $1\bar{1}\bar{1}$ beams. (b),(c) Constructive and destructive interference through 002 and $00\bar{2}$ diffraction discs. (d),(e) Sphalerite space lattice and reciprocal lattice projections for $[110]$ and $[1\bar{1}0]$, indexed for an electron beam traveling down into the page

lowing an $\sim 00\bar{4}$ Kikuchi band. The interaction of the doubly diffracted high-order, odd-index $1\bar{1}9$ and $1\bar{1}\bar{1}$ beams with directly scattered $00\bar{2}$ reflection gives rise to destructive interference through the $00\bar{2}$ disc when all of the reflections are close to the Bragg position. Figure 17.18c, corresponding to a more highly converged incident electron beam, emphasizes the formation of the dark cross through the $00\bar{2}$ diffraction disc. Similarly, Fig. 17.18b accentuates the bright cross through the opposite 002 diffraction disc, formed by constructive interference following interaction with the corresponding $1\bar{1}11$ and $1\bar{1}9$ high-order, odd-index reflections. Figures 17.18d,e, showing the $[110]$ and $[1\bar{1}0]$ projections of the sphalerite space lattice and reciprocal lattice, respectively, are also shown here, to aid understanding of this particular sample geometry. Establishing such microdiffraction patterns requires carefully balancing crystal orientation, probe convergence and layer thickness, and the technique is found to work best with freshly plasma-cleaned sample foils to minimize the effect of extraneous hydrocarbon contamination (which diffuses out the fine-scale contrast under the imaging electron beam).

The dynamical equations of electron diffraction demonstrate that the 002 reflection, which exhibits constructive interference effects (a bright cross), always occurs in the sense of the cation to anion to bond (Fig. 17.18d), and hence is always associated with the sense of advancing $\{\bar{1}\bar{1}\}_b$ planes. To emphasise this, defining $[001]$ as the growth direction of an epilayer on an (001) oriented GaAs substrate, and establishing a bright cross through an 002 diffraction disc, for example through the interaction of $1\bar{1}9$ and $1\bar{1}\bar{1}$ reflections, corresponds to the sense of advancing $(\bar{1}\bar{1})_b$ and $(1\bar{1})_b$ planes in the growth direction, which corresponds to the absolute $[110]$ projection of the sample foil. Conversely, a dark cross through an 002 diffraction disc, such as that obtained through the interaction of $1\bar{1}9$ and $1\bar{1}11$ reflections, corresponds to the sense of advancing $(111)_a$ and $(\bar{1}\bar{1}\bar{1})_a$ planes in the growth direction, which corresponds to the $[1\bar{1}0]$ projection of the sample foil. Since microdiffraction patterns are directly sensitive to crystallographic polarity, a qualitative interpretation of results thus allows the crystal orientation to be unambiguously determined, in situ within the TEM, without reference to any other technique.

Contrast reversals within diffraction discs of systematic row CBED patterns can similarly be used to determine absolute crystal polarity, since Friedel's law again breaks down due to dynamical scattering. In prac-

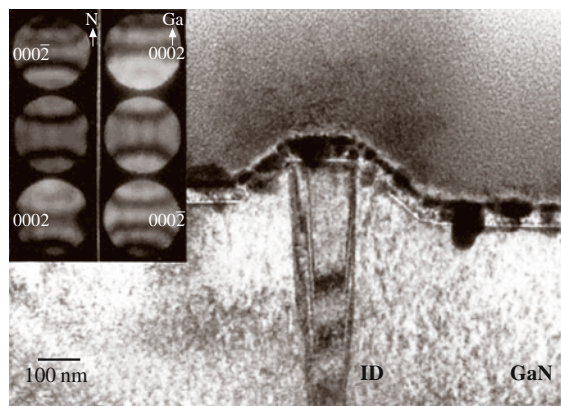


Fig. 17.19 Inversion domain within GaN with CBED patterns recorded across the boundary plane *inset*

tice, this approach is found to be most effective when used with noncentrosymmetric crystals that have much larger differences in anion and cation sizes, such as wurtzite GaN [17.22]. Figure 17.19 corresponds to a thin section through a columnar defect, imaged within homoepitaxial GaN grown on an N-polar $(000\bar{1})$ GaN substrate, with two associated CBED patterns inset, recorded either side of the boundary plane corresponding to matrix and core material, respectively. Simulation of the contrast within the 0002 diffraction discs, for a known sample foil thickness, demonstrates that the central bright and dark bands correspond to N and Ga-polar growth directions, respectively. The reversal of contrast within the 0002 diffraction discs across the boundary plane therefore indicates an inversion in crystal polarity, confirming the defect to be an inversion domain in this instance.

17.4.3 Reflection High-Energy Electron Diffraction

Electron-transparent samples are required for TED investigations. In the case of semiconductor heterostructures, for example, sample preparation requires sequential mechanical polishing and ion beam thinning, which can be very time-consuming, but the advantage of this approach is that diffraction data can be directly correlated with the projected image of the internal structure of the sample. A complementary approach is to use glancing angle electron diffraction techniques to characterize the near-surface microstructure of a bulk sample or an as-grown wafer. Coupling of the electron beam with the material surface at low angle allows scattering of the electrons to occur to produce a diffraction pat-

tern that may be viewed directly on a phosphor screen. This provides valuable information on the near-surface crystallography of the sample, which can be correlated with the growth conditions used or applied surface modifications, without the need for time-consuming sample preparation.

There are three variants of reflection electron diffraction (RED) depending on the accelerating voltage available: low (< 1 kV), medium (1–20 kV) or high (20–200 kV) energy. Medium-energy electron diffraction (MEED) systems are commonly associated with UHV growth chambers, whilst a variant of RHEED can be performed using a conventional TEM. In general terms, coupling of an electron beam with a flat single crystal tends to be associated with the formation of diffraction streaks normal to the surface, providing information on the reconstruction of the atomic layer at the surface. More precisely, streaky RED patterns are indicative of a surface that is not quite perfectly flat, but has slight local misorientations combined with some degree of surface disorder [17.23]. When electrons are coupled with a slightly rougher surface, there is a tendency for more three-dimensional information to be obtained from the interaction with the near-surface microstructure. This leads to the production of spotty diffraction patterns from crystalline materials (half-obscured by the sample shadow edge), in an analogous fashion to TED. It is noted that UHV-MEED systems also allow the intensity fluctuations within the central beam to be monitored in real time, and this provides a way to control layer-by-layer growth.

In practice, RHEED is most effectively used for rapid comparative studies of sets of samples; for example for appraising the effect of changing the growth parameter on the structural integrity of a deposited thin film [17.24]. This approach to process mapping provides a convenient way to identify appropriate samples for more detailed TEM investigation prior to sample foil preparation. There are two variants of the RHEED technique that may be used in a TEM: one where small samples are mounted vertically in the objective lens pole piece, and the other where larger samples are mounted vertically below the projector lens; we focus on the latter variant here.

A schematic diagram of the diffraction geometry for a RHEED experiment is shown in Fig. 17.20a. The diffraction spacing R_{hkl} can be measured and the associated crystal plane spacing d_{hkl} determined using the equation, $\lambda L = R_{hkl} d_{hkl}$, where λL is known as the camera constant. Figures 17.20b,c show a RHEED stage made to interface with a TEM through an existing camera port at the base of the projector lens. The vacuum system of the TEM vents the space below the projector lens upon opening the camera chamber, and so the only waiting time is for the chamber to pump down upon changing a specimen. The RHEED stage is able to support centimeter-square sections of a crystalline specimen, whilst full tilt, twist and lateral movement enables any zone axis within the growth plane of the sample to be accessed. A shadow image of a sample, as projected onto the microscope phosphor screen, is shown in Fig. 17.20d. The area sampled by the glancing elec-

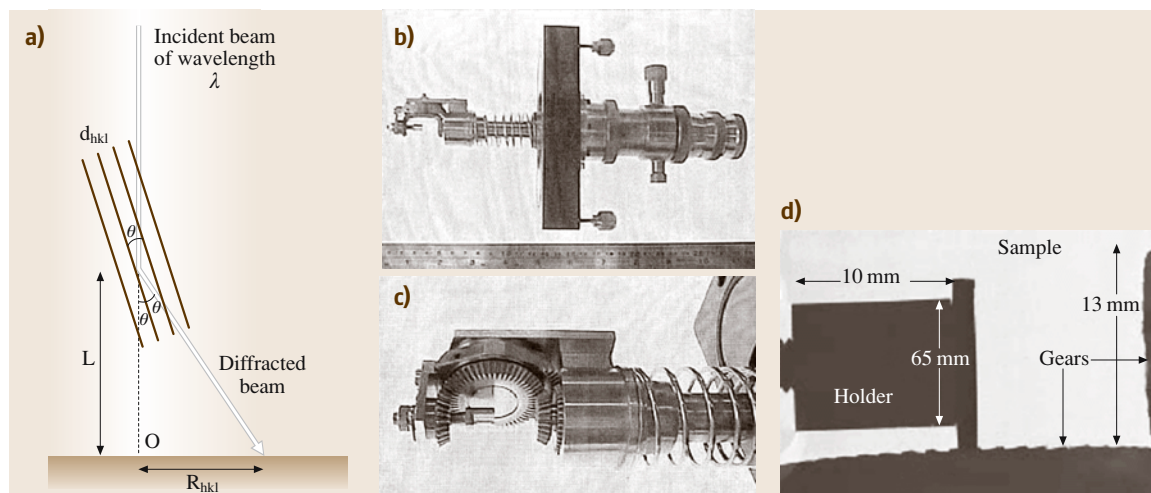


Fig. 17.20 (a) Schematic diagram of the RHEED diffraction geometry; (b) RHEED stage; (c) magnified view of the sample holder; and (d) projected shadow image of a specimen on the microscope phosphor screen

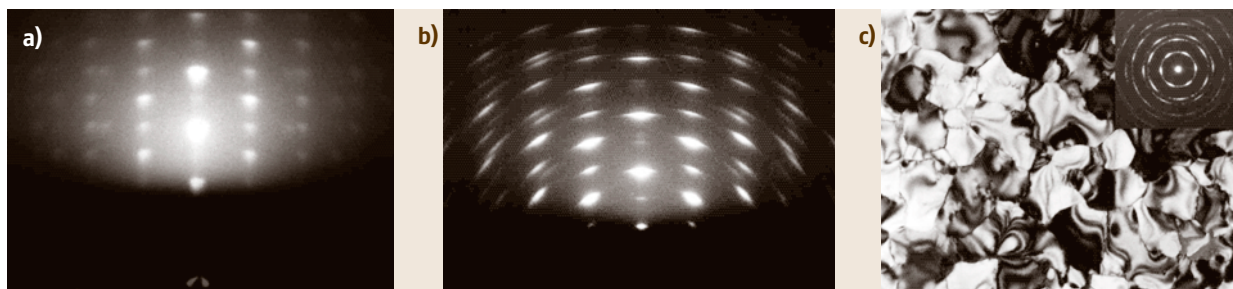


Fig. 17.21a–c RHEED patterns from Si-doped GaN/GaAs(001) (a) before and (b) after plasma cleaning. (c) Plan-view TEM image confirming the presence of rotated columnar grains (TED pattern inset)

tron beam is typically $\approx 1 \text{ mm}^2$, and so this experimental arrangement is closely associated with XRD, although the acquisition time for RHEED data is extremely short, with photographic plate exposure times of $\approx 1 \text{ s}$. Three or four samples can typically be examined within an hour.

The 100 kV RHEED patterns shown in Figs. 17.21a,b were acquired from a highly Si-doped GaN/GaAs(001) heterostructure, grown by MBE at 700°C , before and after plasma cleaning. Surface hydrocarbon deposits arising from specimen handling can generate an amorphous background glow that can hinder RHEED pattern

acquisition (Fig. 17.21a). The oxygen–argon plasma acts to remove such contaminants, allowing high-contrast RHEED patterns to be obtained (Fig. 17.21b). In this instance, the generated RHEED pattern remained effectively constant as the sample was rotated, indicating a random distribution of columnar grains. This predicted microstructure was subsequently confirmed by means of conventional plan-view TEM imaging (Fig. 17.21c, with TED pattern inset), which revealed a fine-scale distribution of rotated columnar grains ($\approx 20^\circ$ of arc).

The RHEED patterns shown in Figs. 17.22a–f, acquired from a variety of III–V heterostructures grown by MBE, are presented to illustrate the variety of microstructures that can be readily distinguished. In the first instance, variations in spot, arc or ring spacings from the central beam provide evidence of the different phases present within a sample and show the general nature of the structural integrity of the near-surface layer (presumed to be representative of the bulk or thin film). Thus, polycrystalline, preferred orientation or single-crystal growth (Fig. 17.22a–c) can be rapidly distinguished, whilst embedded phases (Fig. 17.22d) and anisotropic defect distributions within zinc blende thin films (Fig. 17.22e,f) are also revealed.

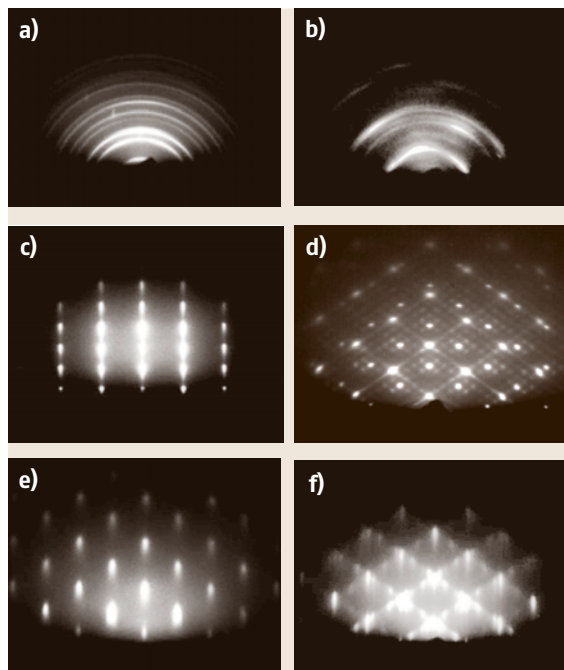


Fig. 17.22a–f RHEED patterns from (a) GaAs/AlN(As)/GaP(001) indicating polycrystalline growth; (b) GaN(As)/GaP(001) grown under low As flux at 620°C showing disordered growth with some degree of preferred orientation; (c) Mg-doped GaN/GaAs{111}B showing α -GaN single-crystal ($\langle 11\bar{2}0 \rangle$ projection); (d) GaN(As)/GaP(001) grown under high As flux at 650°C showing spots due to single-crystal GaAs and β -GaN; (e), (f) Be-doped GaN/GaAs(001) grown at 700°C showing orthogonal $\langle 110 \rangle$ and $\langle 1\bar{1}0 \rangle$ projections. Extra spots indicate a high degree of anisotropy for the distribution of defects within this sample

17.5 Characterizing Functional Activity

There are many solid state analytical techniques that employ X-ray or electron probes, generating a variety of signals for chemical microanalysis. Techniques for performing correlated assessment of the structural and functional performance of a material are perhaps less well-covered in mainstream texts. Accordingly, we now briefly introduce the techniques of scanning transmission electron beam induced conductivity (STEBIC) and TEM-cathodoluminescence (TEM-CL), since these allow us to make correlated structure–property investigations of electrical and optical activity within a semiconductor, respectively.

As discussed earlier, when an electron beam is incident on a semiconductor specimen, electron–hole pairs are created by the excitation of crystal electrons across the band-gap. These electron–hole pairs can, for example, recombine to emit light that may be detected by a photomultiplier. A CL image can then be obtained by displaying the detected photomultiplier signal as a function of the position of the incident electron beam as it is scanned across the specimen. CL spectra can also be acquired in spot mode, which show features attributable to excitons, donor–acceptor pairs or impurities. The ‘information content’ of CL images and spectra therefore includes the location of recombination sites such as dislocations and precipitates, and the presence of doping-level inhomogeneities. Similarly, if the sample is configured to incorporate a collection junction, such as a Schottky-contacted semiconductor or an ohmic-contacted p–n junction, electron–hole pairs that sweep across the built-in electric field constitute current flow. This can be amplified and an image of the recombination activity displayed as the electron beam is rastered

across the sample. If the dislocations within a semiconductor act as nonradiative recombination centers, then they appear as dark lines in both CL and EBIC images because of the reduced specimen luminescence or reduced current that is able to flow through the collection junction when the beam is incident at a defect.

The techniques of CL and EBIC are most commonly performed in an SEM, but this precludes the direct identification of features responsible for a given optical or electronic signature. The resolution of extended defects achieved using EBIC and CL techniques is limited by the penetration depth of the electron beam, the effect of beam spreading and the diffusion length of minority carriers. Conversely, the resolutions of the STEBIC and TEM-CL techniques, as applied to an electron-transparent sample foil, are essentially limited by specimen geometry. The constraint of minority carrier diffusion length is removed due to the close proximity of the sample foil surfaces, and resolution depends on the incident probe size, the width of the electron hole pair generation zone and the recombination velocity at the free surface. For the case of STEBIC, resolution also depends on the defect position relative to the collecting junction. The trade-off is low electrical signal and a degraded signal-to-noise ratio due to the small generation volume and surface recombination effects, in addition to the practicality of contacting and handling thin foils.

Before presenting a number of material characterization case studies based on electron beam techniques, we now discuss the preparation of electron transparent foils that are free from artefacts and suitable for TEM investigation.

17.6 Sample Preparation

We should initially consider whether destructive or non-destructive preparative techniques need to be applied. Some characterization techniques allow samples to be examined with a minimal amount of preparation, provided they are of a form and size that will fit within the apparatus. For example, the crystallography of bulk or powder samples could be directly investigated by XRD, since the penetration depth of energetic X-rays within a sample is on the scale of $\approx 100\ \mu\text{m}$. The surface morphology and near-surface bulk chemistry of a sample can be directly investigated within the SEM, noting the inter-

action volume of electrons [on the scale of $\approx 1\ (\mu\text{m})^3$] associated with the EDX and WDX techniques. It might, however, be necessary to coat insulating samples with a thin layer of carbon or gold prior to SEM investigation to avoid charging effects. Similarly, minimal preparation might only be required before surface assessment using XPS or RHEED, such as cleaning using a degreasing protocol or plasma cleaning. Accordingly, the focus of this section is to introduce the techniques used to prepare samples for TEM investigation, since the requirement is for specimens that are typically submicrometer in

thickness and free of preparation artefacts. For example, a complex sequence of sequential mechanical polishing, dimpling, ion beam thinning and plasma cleaning may be required to produce a pristine semiconductor heterostructure sample, with each stage of the process being designed to minimize artefacts from the previous stage of the preparation process. The idea is to minimize or eliminate artefacts from the preparation process to ensure that the sample being investigated is representative of the starting bulk material. Care is also needed to avoid artefacts that might be introduced through the interaction of the high-energy electron beam with the sample.

In this context, it is interesting to note how TEM sample preparation techniques have developed over the years. Small particles of MgO, produced by igniting the metal and drifting a specimen grid through the smoke, were typical of samples investigated in the 1940s, along with sample replicas made by a dry stripping technique using formvar film. Biological samples fashioned by enzymatic digestion, staining and microincineration were also possible by 1945. Glass and diamond knife microtomes were introduced in the 1950s and used to section soft biological materials. Advances in the controlled preparation of inorganic materials were made upon the introduction of argon ion beam thinning in the late 1960s, which allowed the cross-sectional observation of semiconductor heterostructures when combined with sequential mechanical polishing and dimpling. Significant development work in this area appears throughout the literature from the 1970s. The problem of surface amorphization, introduced by the argon sputtering process, was minimized by adopting low-voltage milling techniques to define the final electron-transparent sample foil. Low stacking fault energy semiconductors, such as II–VI compounds which are easily damaged or InP-based compounds that suffer from In droplet formation with conventional milling techniques, were also successfully prepared for TEM observation using the technique of iodine reactive ion beam etching (RIBE), otherwise known as chemically assisted ion beam etching (CAIBE), developed in the 1980s. The 1990s, however, saw the development of the most effective raft of sample preparation techniques for functional materials and complex semiconductor device structures in particular: tripod polishing, focused ion beam (FIB) milling and plasma cleaning.

The fine adjustment of micrometer supports is the key to tripod polishing that allows the direct mechanical polishing of specimens down to thickness of $< 10\ \mu\text{m}$, using diamond-impregnated polishing cloths on a stable,

high-torque, low-speed polishing wheel. A brief final stage of low-voltage argon ion beam thinning (with liquid nitrogen cooling) then enables electron-transparent sample foils to be defined, free of differential mechanical polishing and shadowing artefacts, whilst minimizing remnant surface amorphization.

FIB instrumentation (now commonly integrated with an SEM) was originally developed for the semiconductor industry as a diagnostic tool for silicon chip fabrication. The application of a focused beam of gallium ions enables sample material to be sputtered away in a controlled fashion to produce an electron-transparent membrane, for example through a specific device within a complex microprocessor. Samples can be periodically observed using SEs generated by gallium ion/material interactions in order to maintain control over and the precision of the sputtering process. In particular, the ability of FIB instruments to prepare site-specific TEM membranes provides a unique opportunity to access the subsurface microstructure of complex device structures

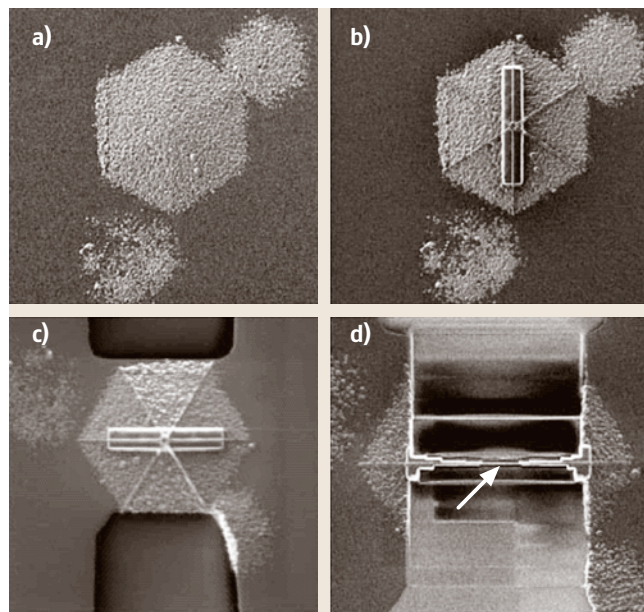


Fig. 17.23a–d SE images showing the application of FIB milling to target a source of defects buried beneath the emergent core of a growth hillock. **(a)** An etched CVD-grown GaN/sapphire hillock ($\approx 5\ \mu\text{m}$ in size). **(b)** Platinum stripes deposited along the hillock facet edges in order to retain sight of the defect core ($\approx 100\ \text{nm}$ in size). **(c)** A high Ga flux is used to create an access trench to the defect. **(d)** Decreasing the Ga flux enables an electron-transparent membrane (arrowed) to be defined at the approximate position of the defect core

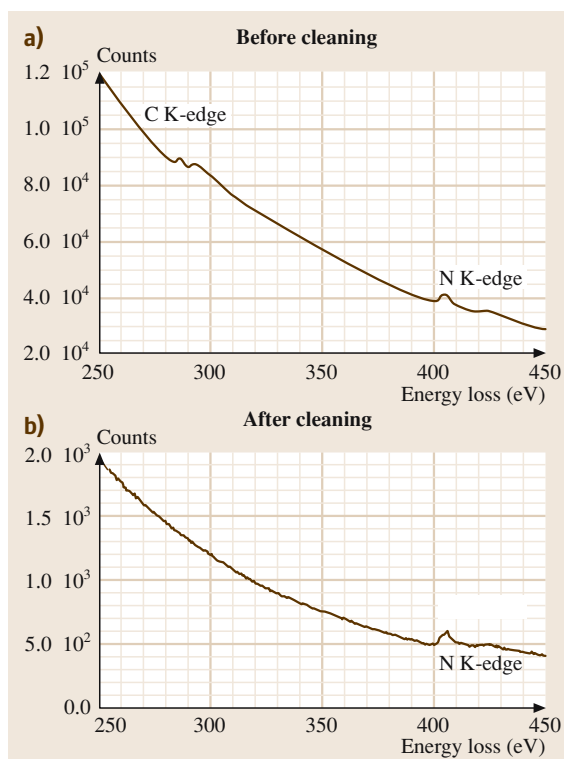


Fig. 17.24a,b EELS spectra from GaN (a) before and (b) after plasma cleaning

or specific regions within a heterostructure identified as being particularly interesting from observations of the surface. Again, by way of example, Fig. 17.23 shows a TEM membrane sectioned through the apex of a hillock identified within a sample of heteroepitaxial GaN–sapphire. Prior to sectioning, the sample was coated with a thin layer of gold and a platinum alkyl decomposed under the rastered Ga beam to de-

fine protective metallic stripes. Sequential sputtering and reduction of the incident Ga beam current allowed a supported thin membrane to be defined (Fig. 17.23d). Care is still required to minimize amorphization artefacts at the surfaces of such membranes, arising from the sputtering action of the glancing, high-energy Ga ion beam. Charging effects associated with insulating samples can act to compromise the fine-scale control of the ion milling procedure. However, this particular problem can be addressed by using low ion dose and shadowing techniques, combined with charge neutralization procedures to inhibit the deflection of the incident ion beam.

Plasma cleaning using oxygen–argon gas, commonly used for the final stage of TEM sample preparation, enables pristine electron-transparent membranes to be obtained, suitable for detailed chemical microanalysis. The disassociated oxygen component of the plasma reacts with organic surface contaminants to produce CO, CO₂ and H₂O reaction products that can be conveniently pumped away. To illustrate the effectiveness of this procedure, it was found that the quality of EELS data from GaN was significantly improved following plasma cleaning. Figure 17.24 shows the removal of the artefact carbon K-edge and enhancement of the sample nitrogen K-edge after a few minutes of exposure to the plasma. It is worth noting that in principle the argon component of the plasma permits gentle sputtering of the sample to occur if low-pressure conditions are used due to the larger mean free path and hence increased energy of the ions. This can be problematic due to the possibility of sample cross-contamination with material sputtered from the supporting sample rod. Another cautionary note on plasma cleaning relates to semi-insulating samples that can become too clean and consequently more susceptible to charging effects under the imaging electron beam.

17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials

There are four general levels of interest when characterizing a given sample:

- What is it made of?
- What additional imperfections does it contain?
- How did it get to be that way?
- How does the microstructure influence the functional properties of the material/device?

For example, one might wish to identify the chemical constituents and crystal structure of a sample to start with, such as whether it is a compound or an alloy, and whether it is single-crystal, polycrystalline, exhibits a preferred orientation or is amorphous. One might then wish to characterize the additional fine-scale defect microstructure within the sample, including precipitates or extended structural defects such as dis-

locations, stacking faults or domain boundaries, since these could deleteriously affect the functional properties of the material. The next level of understanding focuses on how the material has formed via growth, processing or device usage, and seeks to make sense of the process of dynamic evolution, since this potentially enables us to find ways to improve the material in a controlled manner. A clear record of sample history is generally helpful in this context, particularly when trying to identify defect sources. The final level of understanding is possibly the most challenging, since it seeks to associate microstructure with the functional properties exhibited by the sample, and thereby to make sense of its structure–property–processing interrelationship at the fundamental level.

In this context, structural characterizations of semiconductor heterostructures tend to focus on the following issues:

- The integrity of the layer growth and the orientation relationship with the substrate.

- The nature of the structural defects within the epilayer, arising (for example) from the lattice mismatch or from differential thermal contraction following cool-down from the growth temperature.
- The structural integrities of the critical interfaces within the device's active region, such as multiple quantum wells, and the chemical uniformities of the associated alloy layers.
- Modification of the microstructure due to subsequent processing, such as contact formation and device usage.

XRD techniques are well-suited to assessments of the structural integrities of bulk and epitaxial thin film semiconductors. However, we should note the inherent averaging over a large number of microscopic features, such as dislocation distributions responsible for twist and tilt of mosaic grains, associated with such techniques. Conversely, TEM and related techniques are more suited to assessing the fine-scale defect microstructure of a heterostructure. For example, conventional

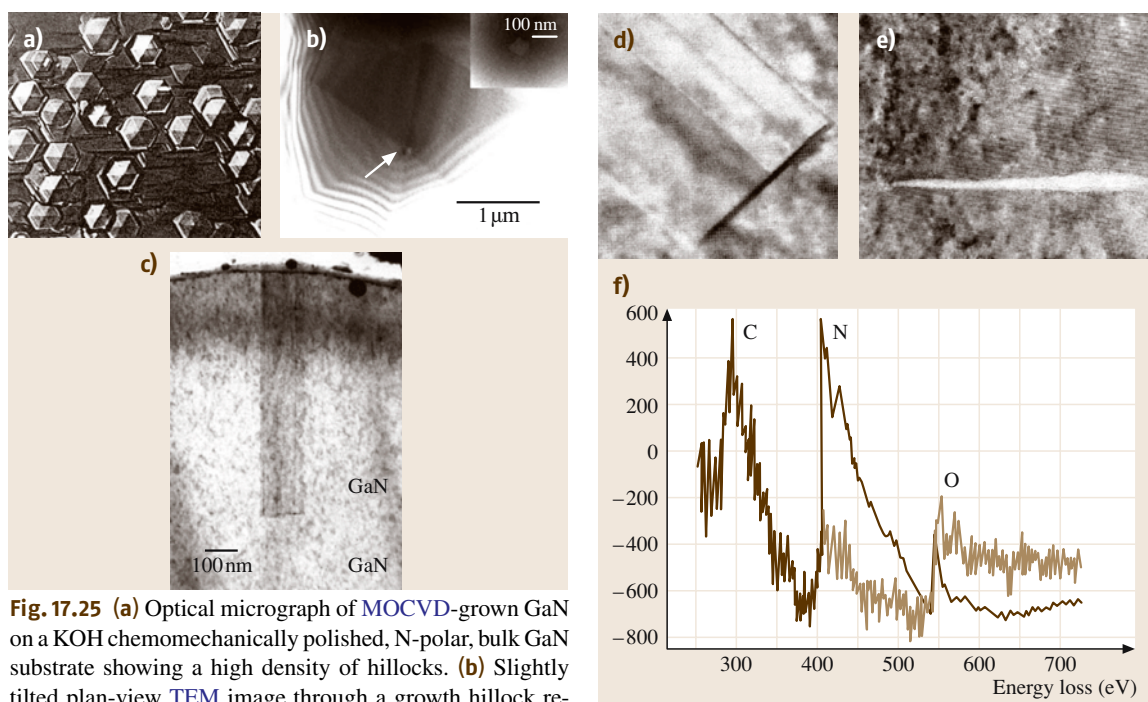


Fig. 17.25 (a) Optical micrograph of MOCVD-grown GaN on a KOH chemomechanically polished, N-polar, bulk GaN substrate showing a high density of hillocks. (b) Slightly tilted plan-view TEM image through a growth hillock revealing a central defect core (arrowed). An enlarged $\langle 0001 \rangle$ projected image of the hillock core is shown in the inset. (c) Cross-sectional, weak-beam TEM image through a hillock core. CBED analysis confirmed the feature to be an inversion domain. (d) HAADF image indicating the

presence of a low atomic number material at the inversion domain source. (e) HREM image indicating the presence of a narrow band of amorphous material at the inversion domain source. (f) EEL second difference spectrum indicating the presence of oxygen at the nucleating event

weak-beam, HREM or CBED analysis can be used for fine-scale defect structural analysis, whilst EELS and EDX analysis can be used to profile alloy composition. However, the ability to perform atomic-level structural characterization and chemical analysis on the nanometer scale is offset by concerns about statistical significance and whether the small volume of material analyzed is truly representative of the larger object. Therefore, electron microscopy-based techniques combined with FIB procedures for site-specific sample preparation tend to be used when investigating integrated device structures.

The examples provided so far illustrate how various diffraction and imaging techniques can provide information on the structural integrity of a given sample. The following examples emphasize the need to apply complementary material characterization techniques in support of the development of semiconductor science and technology.

17.7.1 Identifying Defect Sources Within Homoepitaxial GaN

The emergence of the (In,Ga,Al)N system for short-wavelength light-emitting diodes, laser diodes and high-power field effect transistors has been the semiconductor success story of recent years. In parallel with the rapid commercialization of this technology, nitride-based semiconductors continue to provide fascinating problems to be solved for future technological development. In this context, a study of homoepitaxial GaN, at one time of potential interest for high-power blue-uv lasers, is presented.

The reduction in extended microstructural defects permitted by homoepitaxial growth is considered to be beneficial in the development of nitride-based technology, particularly in view of the evidence confirming that dislocations do indeed exhibit nonradiative recombinative properties. However, in the case of metalorganic chemical vapor deposition (MOCVD)-grown homoepitaxial GaN on chemomechanically polished (000 $\bar{1}$), N-polar substrates, gross hexagonally shaped surface hillocks were found to develop, considered problematic for subsequent device processing [17.25]. The homoepitaxial GaN samples examined in this case study were grown at 1050 °C. The bulk GaN substrate material was grown under a high hydrostatic pressure of nitrogen (15–20 kbar) from liquid Ga at 1600 °C. Prior to growth, the (000 $\bar{1}$) surfaces were mechanically polished using 0.1 μm diamond paste and then chemomechanically polished in an aqueous KOH solution. Epitaxial growth was performed using trimethylgallium and NH_3 precursors

with H_2 as the carrier gas, under a total pressure of 50 mbar. Figure 17.25a shows an optical micrograph of the resultant homoepitaxial GaN/GaN(000 $\bar{1}$) growth hillocks, typically 5–50 μm in size depending on the layer thickness (and therefore the time of growth).

Electron-transparent samples were prepared in plan view using conventional sequential mechanical polishing and argon ion beam thinning procedures applied from the substrate side, whilst cross-sectional samples were prepared using a Ga-source FIB workstation. As shown earlier, the selectivity of the FIB technique enables cross-sections through the emergent cores of the hillocks to be obtained, thereby allowing the nucleation events associated with these features to be isolated and characterized. When prepared in plan-view geometry for TEM observation, each hillock exhibited a small faceted core structure at the center (Fig. 17.25b), but otherwise the layers were generally found to be defect-free. Low-magnification cross-sectional TEM imaging also revealed the presence of faceted column-shaped defects beneath the apices of these growth hillocks (Fig. 17.25c). It was presumed that these features originated at the original epilayer–substrate interface since no other contrast delineating the region of this homoepitaxial interface could be discerned. A reversal of contrast within the 0002 diffraction discs from CBED patterns acquired across the boundary walls of such features (Fig. 17.19) confirmed that they were inversion domains. Thus, the defect cores were identified as having Ga-polar growth surfaces embedded within an N-polar GaN matrix. Once nucleated, the inversion domains exhibited a much higher growth rate than the surrounding matrix, being directly responsible for the development of the “circus tent” hillock structures around them. Competition between the growth and desorption rates of Ga and N-polar surfaces allowed the gross hexagonal pyramids to evolve.

This initial approach of applying electron diffraction and imaging techniques thus enabled the nature of the inversion domains to be identified and their propagation mechanism established in order to explain the development of the hillocks. However, more detailed chemical analysis was required to ascertain the nature of the source of the inversion domains and how this related to the substrate preparation and growth process. A high-angle annular dark field (HAADF) image of the inversion domain nucleation event is shown in Fig. 17.25d. HAADF is a scanned electron probe imaging technique with a resolution defined by the size of the incident probe, while the scattering (and hence con-

trast) is governed by the local average atomic number. In this instance, the sample, tilted slightly to minimize the effects of diffraction contrast, showed dark contrast at the position of the inversion domain source, confirming the presence of a low atomic number material associated with the nucleation event. HREM subsequently confirmed that such nucleation events were due to narrow bands of amorphous material, 2–5 nm in thickness (Fig. 17.25e), whilst EELS confirmed the presence of oxygen (Fig. 17.25f) within these narrow amorphous bands. Accordingly, these defect sources were attributed to remnant contamination from the chemomechanical polishing technique used to prepare the substrates prior to growth. The oxygen-containing residue was presumed to be gallium oxide or hydroxide – probably products of the reaction of KOH etchant with GaN. An improved surface preparation method incorporating a short, final deoxidizing polishing procedure in an aqueous solution of NaCl led to a dramatic reduction in these nucleation sources and thus allowed N-polar homoepitaxial GaN films to be grown virtually free of these gross hillock structures.

17.7.2 Cathodoluminescence/Correlated TEM Investigation of Epitaxial GaN

The CL technique is ideally suited to studies of luminescence uniformity and spectral purity. The following case study illustrates how the defect microstructure of mixed-phase epitaxial GaN/GaAs(111)B can be correlated with the luminescent properties of the layer [17.26].

Even though the majority of developments in GaN technology to date have come from material grown by MOCVD on sapphire and SiC substrates, there is much interest in exploring alternative growth techniques. The MBE technique offers a lower growth temperature than MOCVD and hence enables a greater range of candidate substrate materials to be investigated. One general issue for the MBE growth of heteroepitaxial GaN is the need for direct control of the process of nucleation, since this impacts on the phase and polarity of the deposit and the resultant structural integrity of the film. Epitaxial GaN preferentially adopts the wurtzite phase, with a band-gap of 3.4 eV_{hex}, although zinc blende inclusions, with a band-gap of 3.2 eV_{cubic}, are sometimes associated with MBE-grown material (for conditions of high Ga flux); see Fig. 17.26. The CL spectrum shown in Fig. 17.26a was recorded from a plan-view, electron-transparent foil of nominally single-crystal wurtzite GaN, cooled to liquid nitrogen temperatures. The peak in the CL spectrum at 386 nm (3.21 eV) was used to cre-

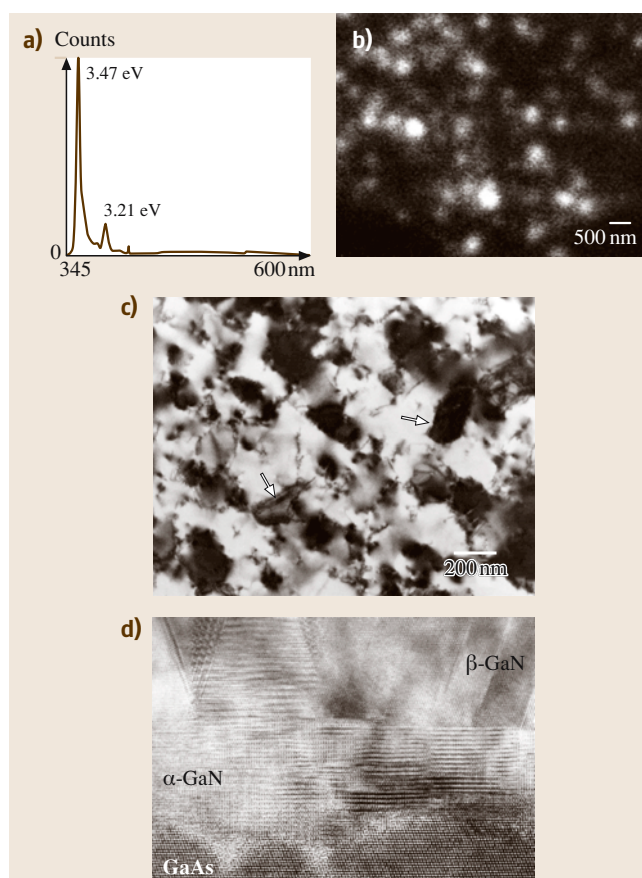


Fig. 17.26 (a) CL spectrum acquired from an electron-transparent foil showing peaks at 357.8 nm and 386 nm, corresponding to wurtzite and zinc blende GaN, respectively; (b) CL image formed at 386 nm and (c) complementary TEM image confirming a distribution of cubic GaN inclusions (arrowed) embedded within the hexagonal GaN matrix; (d) HREM image of epitaxial GaN/GaAs(111)B viewed in cross-section, indicating the nucleation of cubic phase inclusions at the epilayer–substrate interface

ate the image in Fig. 17.26b indicating the distribution of cubic phase inclusions throughout the hexagonal GaN matrix. The complementary conventional TEM image of this plan-view sample foil (Fig. 17.26c) allows the nature of the fine-scale microstructure to be characterized, with diffraction patterns confirming the presence of embedded sphalerite GaN. The size and distribution of these inclusions correlated nicely with the distribution of bright spots in the CL image of 17.26b. A subsequent cross-sectional investigation of GaN/GaAs(111)B specimens confirmed that the cubic inclusions were nucleated at the epilayer/substrate interface (Fig. 17.26d).

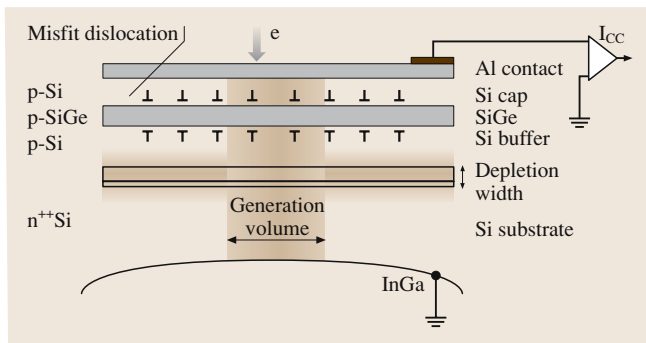


Fig. 17.27 (a) Schematic of the Al/Si/Si_{1-x}Ge_x/Si(001)/InGa sample configuration used for STEBIC investigation.

17.7.3 Scanning Transmission Electron Beam Induced Conductivity of Si/Si_{1-x}Ge_x/Si(001)

The STEBIC technique was originally demonstrated in the late 1970s, using dedicated STEM instrumentation to obtain information on the electrical properties of dislocation core structures within (Ga,Al)(As,P), and thereby providing the first evidence that non-radiative recombination processes at dislocations are related to jogs and kink sites. Dissociated 60° dislocations showed the highest electrical activity, while sessile Lomer-Cottrell edge dislocations were found to be electrically neutral, indicative of reconstructed core structures. STEBIC imaging of an electron-transparent foil allows the electrical and structural properties of defects to be observed simultaneously. The availability of electron sources with high brightness in modern scanning TEM instruments compensates for the main problem of small generation volume and provides an accessible way to perform the STEBIC technique.

The next case study illustrates how the electrical nonradiative recombination properties of MBE-grown Si/Si_{1-x}Ge_x/Si(001) heterostructures correlate with the distribution of interfacial misfit dislocations [17.27]. The Si_{1-x}Ge_x system has potential applications in devices with high electron and hole mobilities. However, the introduction of dislocation networks, or the multiplication of existing dislocations, driven by the 4% misfit strain between Ge and Si, is generally regarded as being detrimental to device operation. With a view to gaining an improved understanding of the relationship between fine-scale structural defects and electron transport properties, structures of capped MBE-grown Si/Si_{1-x}Ge_x/Si(001) were investigated using STEBIC and a range of complementary microscopies. The samples incorporated buried p-n junctions to assist with charge collection and surmount the problem of surface recombination effects.

To prepare for the STEBIC investigation of the Si/Si_{1-x}Ge_x/Si(001), evaporated Al contacts were attached to the top surface prior to preparing electron-transparent foils in plan view by sequential mechanical polishing and argon ion milling of the substrate (Fig. 17.27a). Ohmic contact to the lower surface was made using an InGa eutectic. Electrical activity images were acquired using an electrical contact stage and a scanning TEM. Signal amplification was performed using an amplifier with a low noise current. By controlling the STEM scan rate, the beam could be rastered at a rate compatible with the low bandwidth constraint of the amplifier. STEBIC signals were typically ≈ 100 pA for an electron-transparent Si/Si_{1-x}Ge_x/Si foil imaged at a magnification of $\times 1000$.

Figure 17.27a illustrates the physical parameters relevant to this STEBIC experiment. The silicon substrate was n-type 10^{18} cm^{-3} , whilst the epilayer was p-type boron-doped to a level of 10^{16} cm^{-3} . These

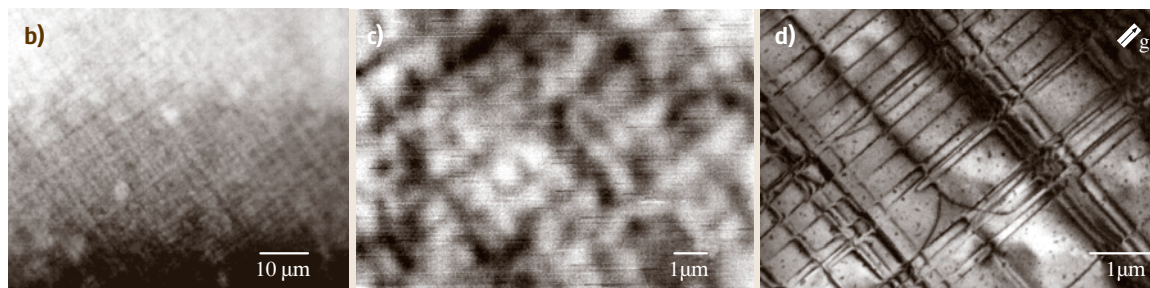


Fig. 17.27 (b) STEBIC image showing recombination activity due to $\langle 110 \rangle$ orthogonal arrays of misfit dislocations within relaxed Si/Si_{0.96}Ge_{0.04}/Si(001). **(c)** Higher magnification STEBIC image illustrating sub- μm resolution of the electrical activity. **(d)** Bright-field TEM image demonstrating the presence of bundles of misfit dislocations ($g = 220$)

values of sample doping were chosen to create a depletion region of ≈ 30 nm width to assist with charge collection, away from the misfit dislocations delineating the $\text{Si}/\text{Si}_{1-x}\text{Ge}_x$ and $\text{Si}_{1-x}\text{Ge}_x/\text{Si}$ interfaces. The low-magnification STEBIC image shown in Fig. 17.27b illustrates the recombination activity within a relaxed $\text{Si}/\text{Si}_{0.96}\text{Ge}_{0.04}/\text{Si}(001)$ sample imaged in plan view. Submicron resolution of the recombination activity is readily achievable using this technique (Fig. 17.27c), with line scans from digitized images indicating a resolution of ≈ 0.3 μm in this case. The spacing of ≈ 1 μm striations in the STEBIC image is much greater than the spacing of individual misfit dislocations shown in the associated TEM image (Fig. 17.27d), and is more closely associated with the spacing of dislocation bundles. Thus, correlation with structural images shows that bunched arrays of orthogonal $\langle 110 \rangle$ misfit dislocations are primarily responsible for the enhanced recombination. For this particular sample, detailed *g.b* analysis confirmed the presence of bands of predominantly 60° misfit dislocations with a few 90° segments arising from dislocation interactions.

Supporting evidence for these dislocations being dissociated and probably decorated by transition metal impurities was obtained from complementary HREM and EDX investigations of metastable and relaxed $\text{Si}/\text{Si}_{1-x}\text{Ge}_x/\text{Si}(001)$ samples from the same growth trial. As the relaxation of a low Ge content, metastable $\text{Si}/\text{Si}_{1-x}\text{Ge}_x/\text{Si}$ structure proceeds, extensive arrays of orthogonal $\langle 110 \rangle$ dislocations form and interact, with dislocations generated in the strained $\text{Si}_{1-x}\text{Ge}_x$ layer being pushed by repulsive dislocation forces into the Si substrate and cap on $\{111\}$ glide planes. HREM observations of relaxed $\text{Si}/\text{Si}_{1-x}\text{Ge}_x/\text{Si}$ samples in cross-section confirmed that the misfit dislocations were dissociated, with tails associated with each of the partials, indicative of impurity decoration (Fig. 17.27e). This particular image was acquired before the development of electron beam-induced damage artefact structures within this sample foil and so is considered representative of the as-grown material. A distribution of small precipitates, of typical size 5 nm, was also identified within metastable samples prior to strain relaxation. These precipitates showed strong scattering in HREM (Fig. 17.27f) and revealed the presence of Fe when analyzed using EDX within a dedicated STEM (1 nm probe size), as shown in Fig. 17.27g. No Fe was present in spectra acquired immediately to the side of the precipitates. (The Ni signal was considered to be an artefact of EDX acquisition and attributed to fluorescence from X-rays and electrons interacting with the specimen's Ni

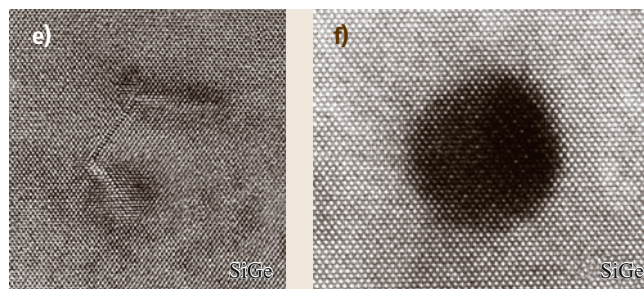


Fig. 17.27 (e) HREM image of a decorated dissociated dislocation viewed in cross-section at the $\text{Si}/\text{Si}_{1-x}\text{Ge}_x$ interface following sample annealing and relaxation. (f) HREM image of a precipitate within as-grown, metastable $\text{Si}/\text{Si}_{1-x}\text{Ge}_x/\text{Si}$.

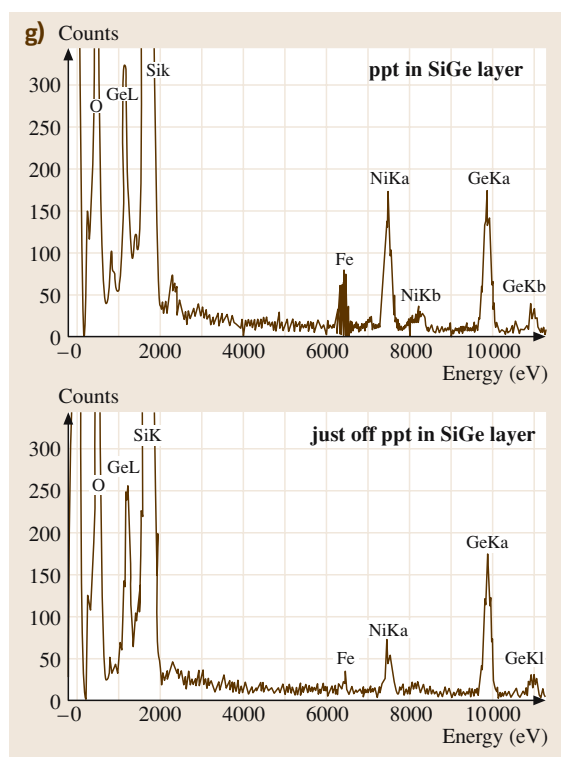


Fig. 17.27 (g) EDX spectra confirming the presence of Fe transition metal impurities

support ring.) Hence, the suggestion is that enhanced transition metal impurity segregation at dissociated dislocations is responsible for the STEBIC contrast observed. This emphasizes the need to control dopants or impurity sources in the vicinity of the heterostructure interface during the development of functional device structures.

17.8 Concluding Remarks

The above commentary has attempted to convey the framework underpinning a variety of analytical techniques used to investigate the structures of semiconductors. It is emphasized that an appropriate combination of assessment techniques should generally be applied, since no single technique of assessment will provide information on the composition, morphology, microstructure and (opto)electronic properties of a given functional material or processed device structure.

This type of considered approach to materials characterization is required in order to break free of the “black-box” mentality that can develop if one is too trusting of the output generated by automated or computerized instrumentation systems. We must always bear in mind the process of signal generation that provides the information content. This in turn should help us to develop an appreciation of performance parameters such as spatial or spectral resolution, in addition to sensitivity, precision and the detection limit. We should consider technique calibration and the appropriate use of standards in order to ensure that the data acquired is appropriate (and reproducible) to the problem being addressed. Consideration should also be given to the form and structure of the data being acquired and how the data sets are analyzed. In this context, distinction should

be made between the processing of analog and digital information and the consequences of data conversion. Issues regarding the interpretation (or misinterpretation) of results often stem from the handling of experimental errors. On a practical level, a rigorous experimental technique should certainly be applied to ensure that the data generated is both meaningful and representative of the sample being investigated, free from artefacts from the preparation and investigation processes. There are clearly differences between qualitative assessment and the more rigorous demands of quantitative analysis. The level of effort invested often reflects the nature of the problem that is being addressed. A comparative assessment of a number of samples may simply require a qualitative investigation (for example, in order to solve a specific materials science problem within a growth or device fabrication process). Alternatively, quantitative analysis may be required to gain a more complete understanding of the nature of a given sample, such as the precise composition. To summarize, an awareness of the methodology used in any investigation is required to establish confidence in the relevance of the results obtained. A range of complementary analysis techniques should ideally be applied to gain a more considered view of a given sample structure.

References

- 17.1 R. W. Cahn, E. Lifshin: *Concise Encyclopedia of Materials Characterization* (Pergamon, New York 1992)
- 17.2 J. M. Cowley: *Electron Diffraction Techniques*, Vol. 1, 2 (Oxford Univ. Press., Oxford 1992, 1993)
- 17.3 B. D. Cullity, S. R. Stock: *Elements of X-Ray Diffraction*, 3rd edn. (Addison Wesley, New York 1978)
- 17.4 J. W. Edington: *Practical Electron Microscopy in Materials Science* (Philips Electron Optics, Eindhoven 1976)
- 17.5 R. F. Egerton: *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Plenum, New York 1996)
- 17.6 P. J. Goodhew, F. J. Humphreys, R. Beanland: *Electron Microscopy and Analysis* (Taylor Francis, New York 2001)
- 17.7 P. J. Grundy, G. A. Jones: *Electron Microscopy in the Study of Materials* (Edward Arnold, London 1976)
- 17.8 P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan: *Electron Microscopy of Thin Crystals* (Butterworths, London 1965)
- 17.9 I. P. Jones: *Chemical Microanalysis Using Electron Beams* (Institute of Materials, London 1992)
- 17.10 D. C. Joy, A. D. Romig, J. I. Goldstein: *Principles of Analytical Electron Microscopy* (Plenum, New York 1986)
- 17.11 M. H. Loretto, R. E. Smallman: *Defect Analysis in Electron Microscopy* (Chapman Hall, London 1975)
- 17.12 D. Shindo, K. Hiraga: *High-Resolution Electron Microscopy for Materials Science* (Springer, Berlin, Heidelberg 1998)
- 17.13 J. C. H. Spence: *Experimental High-Resolution Electron Microscopy – Fundamentals and Applications* (Oxford Univ. Press, New York 1988)
- 17.14 G. Thomas, M. J. Goringe: *Transmission Electron Microscopy of Metals* (Wiley, New York 1979)
- 17.15 D. B. Williams, C. B. Carter: *Transmission Electron Microscopy: A Textbook for Materials Science* (Plenum, New York 1996)
- 17.16 R. Hull, J. C. Bean: *Crit. Rev. Solid State* **17**, 507 (1992)
- 17.17 T. Sugahara, H. Sato, M. Hao, Y. Naoi, S. Kurai, S. Tattori, K. Yamashita, K. Nishino, L. T. Romano, S. Sakai: *Jpn. J. Appl. Phys.* **37**, 398 (1997)
- 17.18 Y. Xin, P. D. Brown, T. S. Cheng, C. T. Foxon, C. J. Humphreys: *Inst. Phys. Conf. Ser.* **157**, 95 (1997)

- 17.19 Y. Ishida, H. Ishida, K. Kohra, H. Ichinose: *Philos. Mag. A* **42**, 453 (1980)
- 17.20 D. B. Holt: *J. Mater. Sci.* **23**, 1131 (1988)
- 17.21 K. Ishizuka, J. Taftø: *Acta Cryst. B* **40**, 332 (1984)
- 17.22 D. Cherns, W. T. Young, M. Saunders, J. W. Steeds, F. A. Ponce, S. Nakamura: *Philos. Mag.* **A77**, 273 (1998)
- 17.23 J. M. Cowley: *Electron Diffraction: An Introduction*, Vol. 1 (Oxford Univ. Press, Oxford 1992)
- 17.24 G. J. Russell: *Prog. Cryst. Growth Ch.* **5**, 291 (1982)
- 17.25 J. L. Weyher, P. D. Brown, A. R. A. Zauner, S. Muller, C. B. Boothroyd, D. T. Foord, P. R. Hageman, C. J. Humphreys, P. K. Larsen, I. Grzegory, S. Porowski: *J. Cryst. Growth* **204**, 419 (1999)
- 17.26 P. D. Brown, D. M. Tricker, C. J. Humphreys, T. S. Cheng, C. T. Foxon, D. Evans, S. Galloway, J. Brock: *Mater. Res. Soc. Symp. Proc.* **482**, 399 (1998)
- 17.27 P. D. Brown, C. J. Humphreys: *J. Appl. Phys.* **80**, 2527 (1996)

Surface Chem

18. Surface Chemical Analysis

The physical bases of surface chemical analysis techniques are described in the context of semiconductor analysis. Particular emphasis is placed on the **SIMS** (secondary ion mass spectrometry) technique, as this is one of the more useful tools for routine semiconductor characterization. The practical application of these methods is addressed in preference to describing the frontiers of current research.

18.1 Electron Spectroscopy	373
18.1.1 Auger Electron Spectroscopy	373
18.1.2 X-Ray Photoelectron Spectroscopy (XPS)	375
18.2 Glow-Discharge Spectroscopies (GDOES and GDMS)	376
18.3 Secondary Ion Mass Spectrometry (SIMS)	377
18.4 Conclusion	384

Surface chemical analysis is a term that is applied to a range of analytical techniques that are used to determine the elements and molecules present in the outer layers of solid samples. In most cases, these techniques can also be used to probe the depth distributions of species below the outermost surface. In 1992 the International Standards Organisation (ISO) established a technical committee on surface chemical analysis (ISO TC 201) to harmonize methods and procedures in surface chemical analysis. ISO TC 201 has a number of subcommittees that deal with different surface chemical analytical techniques and this chapter will discuss the applications of these different methods, defined by ISO TC 201, in the context of semiconductor analyses. In particular, this discussion is intended to deal with practical issues concerning the application of surface chemical analysis to routine measurement rather than to the frontiers of current research. Standards relating to surface chemical analysis developed by the ISO TC201 committee can be found on the ISO TC201 web site www.iso.org (under “standards development”).

Traditional surface chemical analysis techniques include the electron spectroscopy-based methods Auger electron spectroscopy (**AES** or simply Auger) and X-ray photoelectron spectroscopy (**XPS**, once also known as ESCA – electron spectroscopy for chemical analysis), and the mass spectrometry method **SIMS** (secondary

ion mass spectrometry). The ISO TC 201 committee also has a subcommittee that deals with glow discharge spectroscopies. Whilst these latter methods have been used more for bulk analysis than surface analysis, the information they produce comes from the surface of the sample as that surface moves into the sample, and so they have been finding applications in depth profiling studies.

One thing that is common to all of these surface chemical analysis techniques is that they are vacuum-based methods. In other words, the sample has to be loaded into a high or ultrahigh vacuum system for the analysis to be carried out. With the one exception of glow discharge optical emission spectroscopy (**GDOES**), where the analysis relies upon the detection of photons, all of the techniques also depend upon the detection of charged particles. This requirement for vacuum operation necessarily imposes limits on the types and sizes of samples that can be analyzed, although of course instruments capable of handling semiconductor wafers do exist. The quality of the vacuum environment around the sample can also affect the quality of the analysis, especially with regard to the detection of elements that exist in the atmosphere around us. The size and complexity of surface chemical analysis equipment has arguably tended to limit the wider use of these powerful methods.

18.1 Electron Spectroscopy

In the electron spectroscopies, Auger and **XPS**, the surface of the sample is probed by an exciting beam which

causes electrons to be ejected from the atoms in the sample. These electrons are collected and their ener-

gies analyzed. The two techniques are similar but subtly different.

18.1.1 Auger Electron Spectroscopy

In Auger, a beam of electrons is used to excite the sample. During the interaction of the primary electron beam with the sample atoms, core electrons are knocked out, creating vacancies in the inner electron shells. Electrons from outer shells can fall into the vacancy, thus leaving the atom in an unstable state and, in order to return to equilibrium, the excess energy the atom possesses can be dissipated in one of two ways: either by the emission of an X-ray photon with a characteristic energy (the basis of energy- or wavelength-dispersive X-ray analysis), or by the emission of a third electron (the Auger electron), with an energy determined by the difference in energy between the original core state and those of the two other levels involved. Clearly these energies are uniquely determined by the energy levels in the atom and thus provide a route for analysis. As three electrons are necessary for the Auger process to occur, AES is incapable of detecting hydrogen or helium, but all other elements produce characteristic Auger electrons. The energies of the Auger electrons can range from a few tens of electron volts to a few thousand electron volts. In this energy regime, electrons can only travel of the order of monolayers through a solid before an interaction occurs which causes a loss of energy, thus destroying the analytical information the electron possessed. It is this property which gives the Auger technique its surface sensitivity – although Auger electrons will be produced as far into

the material as the primary beam can penetrate, only those produced in the top few monolayers can escape with their characteristic energy intact. The Auger electrons excited deeper into the sample lose energy before escaping and contribute to a background signal, as do the scattered primary electrons and the initial core electrons ejected at the start of the process. These scattered electrons produce a background signal against which the Auger electrons must be detected; this limits the analytical sensitivity that can be achieved by Auger. This is also why Auger spectra are sometimes displayed as the differential of the number of electrons against energy, because the small Auger peaks are more rapidly varying functions than the larger, slowly changing background signal and hence are enhanced by the differentiation process. Figure 18.1 shows an Auger spectrum from GaInAsP, where the spectrum plotted as the number of electrons as a function of energy is compared to the differential of the number of electrons as a function of energy.

As the number of Auger electrons produced is proportional to the number of atoms excited, Auger offers the ability to perform quantitative as well as qualitative analyses, although some form of calibration is required, either through the use of local reference materials or instrument calibration and standard databases. So Auger offers quantitative analysis for all elements from lithium to uranium from layers only a few atoms in thickness. What gives Auger an extra dimension is the ability to profile into the sample by removing the outermost surface layers with an inert gas ion beam (usually argon) in order to expose the layers below. This sputter depth profiling is the inverse of the sputter deposition widely used to deposit thin layers of material. Argon is the most widely used ion beam in Auger depth profiling, as its effects are physical rather than chemical, although it should be noted that some chemicals can be modified by the sputtering process. For example, some metal oxides can be reduced by sputtering while others are not, and so any chemical state data inferred from atomic compositions in sputter depth profiles should be treated with caution. Physical effects can also occur, for example atomic mixing and the development of surface topography, which can distort the shape of buried features.

One of the great advantages of Auger is that, as the excitation is provided by electrons, the primary beam can be easily focused and scanned over the surface of the sample. By detecting the scattered electrons, a physical image of the sample is produced as in the scanning electron microscope, and maps of elemental distributions can be obtained by detecting the Auger signal as

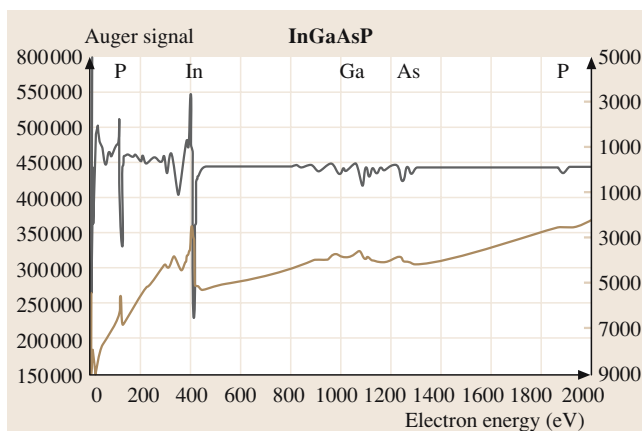


Fig. 18.1 Auger spectrum from GaInAsP, shown as the number of electrons and the differential of the number of electrons as a function of energy

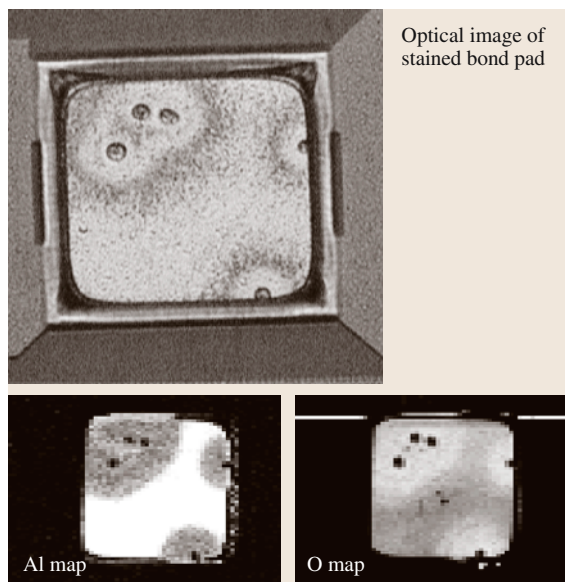


Fig. 18.2 Optical image and Auger maps of a stained aluminium bond pad

a function of beam position. These elemental maps can be time-consuming to acquire but can be useful when making a point or illustrating a book chapter; however, for practical analyses it is often adequate to identify the feature of interest from the physical image. Figure 18.2 shows aluminium and oxygen Auger maps from a contaminated bond pad. In the region of the contamination, the aluminium signal is reduced and the oxygen signal is higher compared with the uncontaminated regions.

In terms of hardware, Auger systems can be stand-alone systems comprising an electron beam column, an electron energy analyzer, an inert gas ion gun and sample handling stage and an associated vacuum chamber (or chambers), or form a part of a multitechnique system with X-ray sources for XPS analysis or a mass spectrometer for basic SIMS studies.

18.1.2 X-Ray Photoelectron Spectroscopy (XPS)

XPS is very similar to Auger in terms of the instrumentation and physics involved. The primary excitation, as the name implies, is, in this case, a beam of X-rays. The X-rays, often magnesium or aluminium $K\alpha$, eject core electrons from the surface atoms by the photoelectron effect. The kinetic energy of the emitted photoelectrons will be equal to the difference between the X-ray photon energy and the core level binding energy, and thus

will be less than ≈ 1100 eV or ≈ 1400 eV for magnesium and aluminium $K\alpha$ respectively. So, as with Auger, the mean free paths of the photoelectrons are of the order of monolayers in solid materials. As a core level vacancy is produced by the excitation, Auger electrons will also be present in the XPS spectra but, by convention, Auger spectroscopy refers to the electron-excited situation. Whereas in Auger electron spectroscopy the electron energy is usually referred to in terms of the electron's kinetic energy, in XPS the electron binding energy is usually plotted as the ordinate in the spectra. This means that, whatever X-ray excitation is used, be it magnesium, aluminium or a more exotic material, the photoelectrons will appear at the same binding energy in the spectra but the apparent positions of the Auger peaks will change (on the binding energy scale) as their kinetic energy is independent of the excitation source. There are no scattered primary electrons (which are always present in Auger spectra) in the XPS spectra, so these spectra have better signal-to-background, and the photoelectron peaks are easily distinguished against the background arising from scattered photoelectrons produced deeper into the sample and other secondary processes.

As the problem of focusing electron beams is much simpler than focusing X-ray beams, XPS is perceived as a large-area technique, whereas Auger is the technique of choice for small area analysis. However, the relentless advances made in the performance and design of instruments means that XPS instruments can achieve spatial resolutions of the order of 1 to $10\mu\text{m}$. In XPS no charge is brought to the sample by the primary excitation, and so insulating samples are easier to analyze with XPS than with Auger; also, the photoelectron peaks show small energy differences in the peaks positions depending upon the local chemical environment of the atom from which they originated, the so-called chemical shift. While chemical shifts are present in some Auger peaks, this is the exception rather than the rule, and XPS is the technique of choice where information on the local chemical state of the surface is required. The physics of the XPS process is probably even better understood than the Auger process, and quantification of the spectra is relatively routine.

As with Auger, composition depth profiles can be produced by sputtering the surface of the sample with an inert gas ion beam. Again caution is advised when interpreting chemical state information from a surface that has been subject to ion bombardment. With both of the electron spectroscopies it is the surface of the sample that remains after sputtering that is analyzed in a depth profile, and there are two factors to be aware

of (if not more). Once the passivating surface layer has been sputtered away, the surface of the sample may become chemically active and getter residual gas from the vacuum system. If the sample is a multicomponent material, one component may have a higher sputtering rate than the other, so as the sputtering process proceeds the surface will become depleted in the higher sputtering rate material. This process will continue until an equilibrium state is reached where the material leaving the surface is in the same ratio as the bulk composition; the

corollary of this is that the surface will be enriched in the lower sputter rate material and so the composition of the material as measured by either *XPS* or Auger will be in error unless this effect is understood and accounted for.

Both Auger and *XPS* are capable of detecting all elements from lithium to uranium (and beyond), and have sensitivities in the parts per hundred to parts per thousand regime. The responses vary from element to element but typically sensitivities remain within an order of magnitude or so between elements.

18.2 Glow-Discharge Spectroscopies (GDOES and GDMS)

These are two apparently similar but quite unrelated techniques that rely upon glow discharges as the excitation source. In glow discharge optical emission spectroscopy (*GDOES*) a high-pressure glow discharge is used to sputter material from the surface of a sample, and this sputtered material is detected by the optical emissions it produces in the glow discharge. In glow discharge mass spectrometry (*GDMS*), a low pressure dc glow discharge is used to sputter material from the sample surface and ionized material from the discharge is extracted into a mass spectrometer for analysis.

GDOES is probably the simplest of all the surface chemical analysis techniques, at least as far as the vacuum requirements are concerned. There is no complex vacuum system, as needed for all of the other methods, and the sample itself sits with atmospheric pressure on one side of it while the opposite face acts as one

electrode of a glow discharge cell that is pumped by a simple vacuum pump. A flow of high-purity argon gas flows through the cell, providing the sputtering and discharge gas and purging the cell of impurities and material removed from the sample. A window at the other end of the discharge cell transmits light from the discharge into an optical spectrometer. By using a spectrometer with a number of photomultiplier detectors, prepositioned at the known wavelengths of the expected element emission lines, data from a large number of elemental channels can be collected in parallel, making *GDOES* an extremely efficient analytical system. A scanning spectrometer can also be included in the instrument to provide a continuous spectral scan to detect emission lines from elements other than those built into the instrument, but this is, of course, a serial detection device and the advantages of parallel acquisition are lost.

Traditionally *GDOES* has been widely used for the analysis of metals and coatings on metals, but it is currently also finding application in the area of semiconductor materials. With the development of radio frequency glow discharge sources, the technique is capable of analyzing insulators, and with the high sputtering rates available, it can depth profile tens of μm into dielectric layers. With fast electronics and parallel detection, thin oxide layers can also be profiled. The sensitivity of the *GDOES* technique lies between that of Auger or *XPS* and that of *SIMS*. One potential weakness of the method is that the technique has no spatial resolution and the analysis area is millimeters in diameter. It is thus useful for large-area plain samples, but cannot be used with patterned material or to probe small features. The technique is useful for bulk analysis, but it is the ability to depth-profile into material, providing an insight into layer structures, that is its greatest appeal. However, because the technique has no ability to dis-

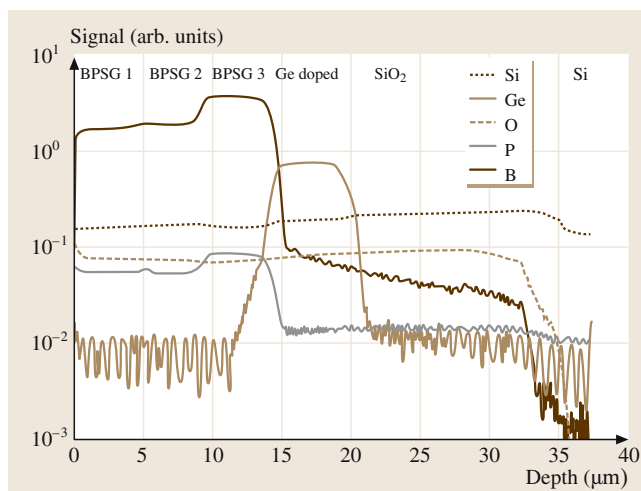


Fig. 18.3 *GDOES* depth profile through a *DWDM* multilayer glass structure

criminate where the analytical signal is coming from, the quality of the depth profiles produced will be compromised by crater edge effects. In other words, while most of the analytical signal will originate from the bottom of the sputtered crater, there will always be some information that comes from the crater side wall. The consequence of this is that, with layered structures, layers closer to the surface will appear to tail into layers beneath them, even though the interface between the layers is abrupt. This effect can be seen in the depth profile shown in Fig. 18.3, which shows a GDOES profile into a DWDM structure.

Glow discharge mass spectrometry (GDMS) is a considerably more complex technique, at least from an instrumental point of view. Originally developed as a method of bulk analysis, GDMS is probably the most sensitive, in terms of the detection limit achievable, of all of the techniques being considered here. As with GDOES, in GDMS the sample forms one electrode in a simple glow discharge cell. However, in the case of GDMS, the discharge cell is mounted within a high-vacuum system. In its original form, the sample (typically be 1 mm² by about 15 mm long) is placed in the center of a cylindrical cell into which argon is leaked at low pressure. By applying a dc voltage between the sample and the cell, an argon plasma is created which

sputters the outside of the sample, removing material. This material, some of which is ionized but the majority of which is neutral as it leaves the surface, is ionized by a variety of processes as it passes through the glow discharge plasma. These ions are then accelerated into a high-resolution magnetic sector mass spectrometer where they are mass-analyzed and counted. Instruments can also be based on quadrupole mass spectrometers, but it is the magnetic sector instruments which offer the greater sensitivity. By sweeping the mass spectrometer through a range of masses, which can cover the entire periodic table, the major, minor and trace elements present in a sample can be determined. GDMS is a particularly powerful method of detecting the trace elements present in bulk semiconductor materials at levels down to parts per billion.

It is also possible to analyze flat, rather than matchstick-shaped, samples in GDMS. Just as in GDOES, the flat sample is positioned at the end of the discharge cell, and a cylindrical crater is etched into the sample surface. As with GDOES, with GDMS there is no spatial resolution, and the depth information from layered structures will be distorted by crater edge effects and loss of crater base flatness as it is not possible to discriminate between ions produced from the base of the crater and those produced from the sidewalls.

18.3 Secondary Ion Mass Spectrometry (SIMS)

SIMS is probably the most powerful and versatile of all of the surface analysis techniques and comes in the widest variety of instrumentations, from big, stand-alone instruments to bench-top instruments and add-ons to electron spectrometers. SIMS can offer chemical identification of submonolayer organic contamination, measurement of dopant concentrations, and can produce maps and depth profile distributions from nanometers to tens of μm in depth. However, no one instrument is going to be capable of all of these tasks, and even if it could it would not be able to achieve all of them at the same time.

SIMS, in its simplest form, requires an ion gun and a mass spectrometer. The sample is placed in a vacuum chamber and ions from the ion gun sputter the sample surface. Material is sputtered from the sample surface and some of this will be ionized, although in most cases the major part of the sputtered material will be in the form of a neutral species. The ionized component of the sputtered material is mass-analyzed with the mass spectrometer.

The technique has evolved in various directions from this common origin to produce a variety of subtly different variants of the SIMS technique, including dynamic SIMS (DSIMS), static SIMS (SSIMS) and time of flight SIMS (ToFSIMS), each of which has its own distinct attributes. There are three main types of mass spectrometer used for SIMS analysis: the magnetic sector, the quadrupole and the time of flight, ToF. Dedicated depth-profiling SIMS machines, dynamic SIMS instruments, tend to employ either magnetic sector or quadrupole mass spectrometers. Magnetic sector instruments offer high transmission and high mass resolution capabilities, useful for separating adjacent mass peaks with a very small mass difference, for example ³¹P from ³⁰SiH. Quadrupole mass spectrometers offer ultrahigh vacuum compatibility and, as well as being used in DSIMS instruments, smaller versions are also found as add-ons to Auger/XPS instruments and bench-top instruments. Time of flight instruments are remarkably efficient in their use of material in that the entire mass spectrum is sampled in parallel, whereas in the

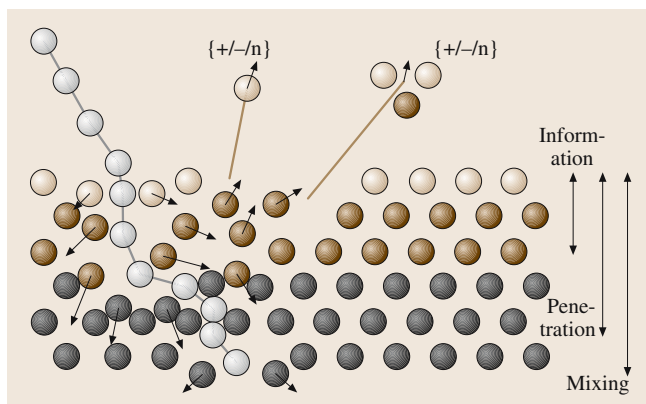


Fig. 18.4 Schematic diagram of the SIMS process

magnetic sector and quadrupole instruments the spectrum is produced by sequentially scanning through the mass range of interest. ToFs can also offer high mass resolution but profile relatively slowly because it is necessary to use a pulsed ion beam with a low duty cycle. Whilst static SIMS can be carried out on either a quadrupole or magnetic sector instrument, the ToF-based instruments are more suited to the task as less material is used in the course of the analysis. However, when dynamic SIMS, where only a few elements need to be monitored, is the goal, the magnetic sector or quadrupole instrument is the appropriate choice, although ToF machines can be used for shallow profiling.

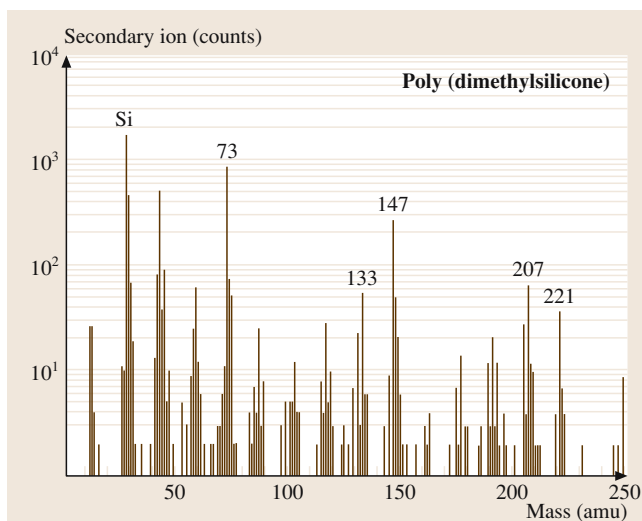


Fig. 18.5 Static SIMS spectrum from a thin film of poly(dimethylsilicone)

Figure 18.4 shows a schematic diagram of the basis of these SIMS techniques. Ions from the ion gun bombard the sample surface and transfer momentum to the atoms in the sample creating a collision cascade. This cascade distributes the energy of the incoming ion amongst the atoms in the sample, causing them to be displaced from their original sites, possibly breaking some bonds between atoms whilst creating others. The bombarding ion also can become implanted into the target material, modifying the chemistry of the material in the process. None of this is peculiar to SIMS – it happens in any sputtering process, be it in Auger depth-profiling, GDMS analysis or sputter deposition systems. It is probable that some of the energy deposited in the sample will cause atoms or molecules to be ejected from the surface of the sample, and those atoms or molecules that leave the surface as ions can be collected and detected by the mass spectrometer.

In the early stages of sputtering, the atoms and molecules sputtered from the surface will originate from areas of virgin surface – in other words from sites that have not yet been damaged by the primary ion – and thus carry with them information about the chemistry of the outer molecular layers of the sample. As the sputtering process proceeds, the probability of the sputtered particles being emitted from an area that has been modified by earlier ion impacts increases. Thus, at the start of the process the sputtered particles are characteristic of the virgin surface, but they will eventually become characteristic of the ion beam-modified surface.

Static SIMS is concerned with the measurement of the sputtered molecules produced at the start of the process, where information about the surface chemistry of the sample can be obtained. In SSIMS a very low dose of primary ions is used, typically less than 10^{13} primary ions per square centimeter, so that there is a very low probability that the ions that are detected come from damaged material. Used in this way, SIMS can be used to identify organic contamination on surfaces and obtain information about the molecular structure of the sample. Figure 18.5 shows a typical SSIMS mass spectrum from poly(dimethyl silicone), a common surface contaminant, which can be recognized by prominent peaks in the mass spectrum at masses 73, 133, 147, 207 and 221, which originate from fragmentation of the parent molecule ($\text{CH}_3 - \text{Si}(\text{CH}_3)_2 - \text{O} - \text{Si}(\text{CH}_3)_2 - \text{O} - \dots - \text{Si}(\text{CH}_3)_2 - \text{CH}_3$).

Whilst the abundance of molecular fragments that can be produced by the sputtering process is of great value for revealing chemical information about the

sample surface in the SSIMS context, it can also be something of a problem when carrying out elemental analysis of trace impurities in simple matrices. For example, the spectrum shown in Fig. 18.6a, from an unknown Ga-based material, illustrates the number and complexity of molecular species produced by the SIMS process, which can make interpretation difficult. Fortunately, the distribution of the number of secondary ions with energy is different between atomic ions and molecular ions. The atomic ions, in general, have a broader energy distribution than the molecular ions, and its energy distribution tends to become sharper as the complexity of the molecular ion increases. For example, Fig. 18.7 shows the ion energy distributions of the Si^+ , Si_2^+ and Si_3^+ ions, illustrating the narrower energy distributions of the molecular ions. Thus, by selecting the ion energy range from which the spectrum is recorded, the molecular information can be suppressed, allowing the elements present in the material to be identified, see Fig. 18.6b.

The intensities of the peaks in the SIMS mass spectra reveal more about the relative ionization probability and instrument transmission function than the number of species of that mass on the sample surface. The signals that are measured in SIMS are proportional to the numbers of ions produced, and this is not simply related to the amount of material present. The degree of ionization can vary enormously from element to element and matrix to matrix. For example, the numbers of positive ions formed by the inert gas bombardment of clean metal samples can be several orders of magnitude lower than the numbers produced from oxidized surfaces of the same metals under identical bombardment conditions. Clearly it is not the concentration of the metal atoms on the surface that is important, as there will be fewer metal atoms in the oxide layer than in the pure metal. It is the presence of oxygen that increases the ionization probability for material leaving the surface as positive ions.

In SSIMS, the chemical nature of the primary ion species is of relatively little consequence in terms of its effect on the number of ions produced by the surface, but in DSIMS, where it is the deliberately modified surface that is of interest, oxygen ion beams are widely used for analyses of electropositive species. Figure 18.8 shows how the silicon matrix signal from a silicon wafer with a native oxide layer behaves as the sputtering process proceeds. Initially there is a strong signal from silicon as the surface oxide layer is still present on the sample surface. The signal then falls as the oxide layer is sputtered away, and then slowly recovers in intensity as

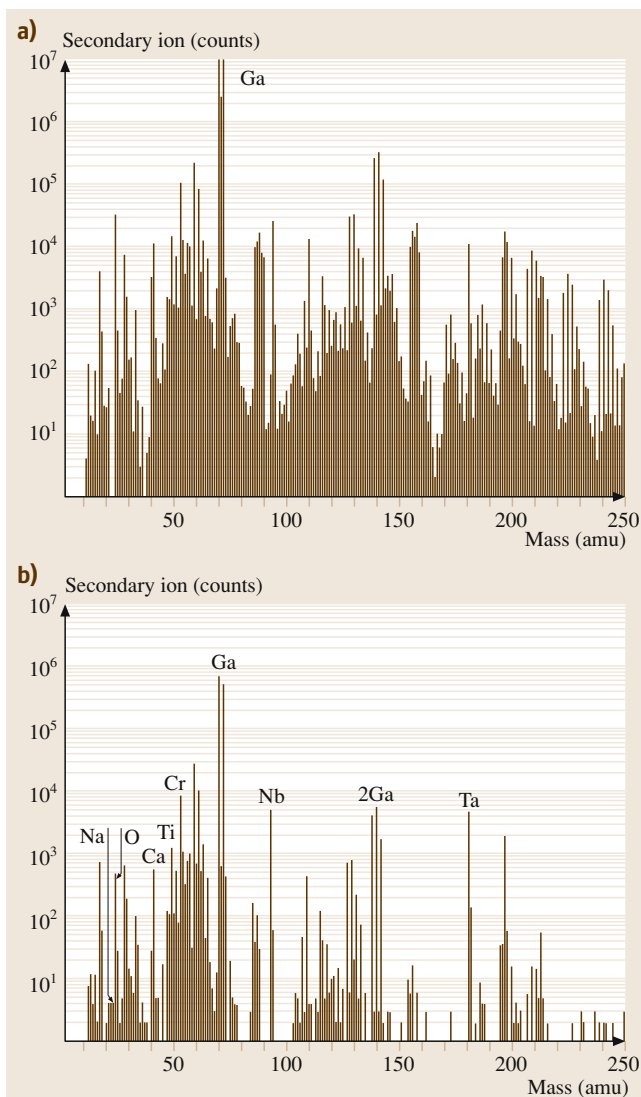


Fig. 18.6 SIMS mass spectrum from an unknown sample *top panel* without energy filtering, *bottom panel* with energy filtering

oxygen from the primary ion beam is implanted into the silicon surface. Eventually a steady state is reached where the material that is being sputtered is constant-composition oxygen-implanted silicon. The thickness of this transient region will depend upon the energy and angle of incidence of the primary oxygen ion beam. The higher the ion beam energy, the thicker the transient region. In the transient region the sputtering rate of the material and the ionization probability may well be changing, and precise quantification in this part of the

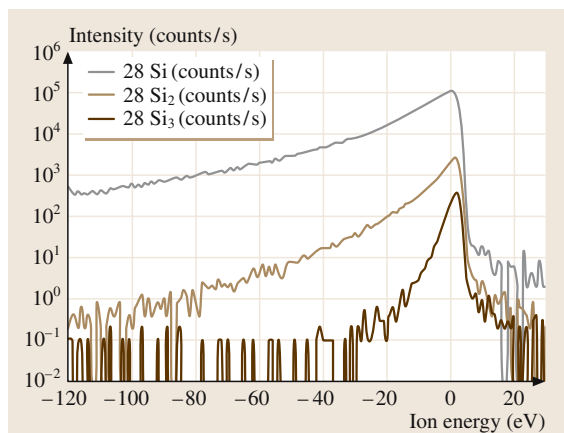


Fig. 18.7 The ion energy distributions of Si^+ , Si_2^+ and Si_3^+ ions

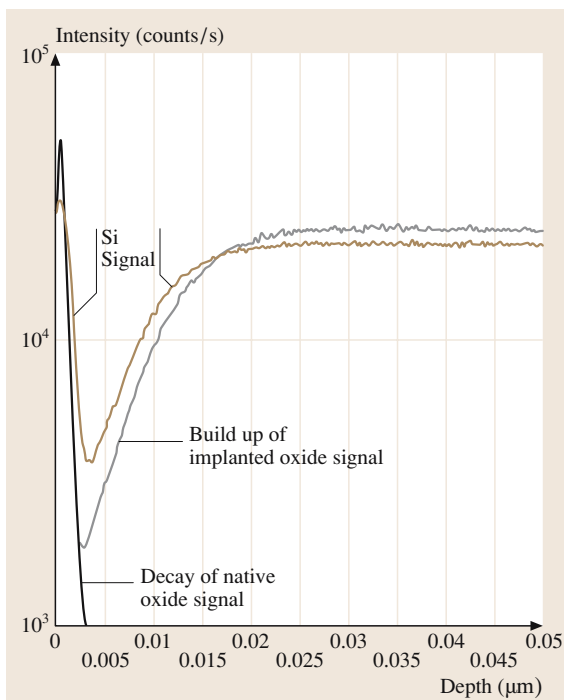


Fig. 18.8 The evolution of the Si^+ signal from a clean silicon wafer as a function of sputtering time with an oxygen primary ion beam, showing the decay of the surface oxide signal and the build-up of the implanted oxide layer

profile may be difficult. This is of particular importance when the features of interest are very close to the surface and within the transient region. The technological importance of shallow structures has driven the develop-

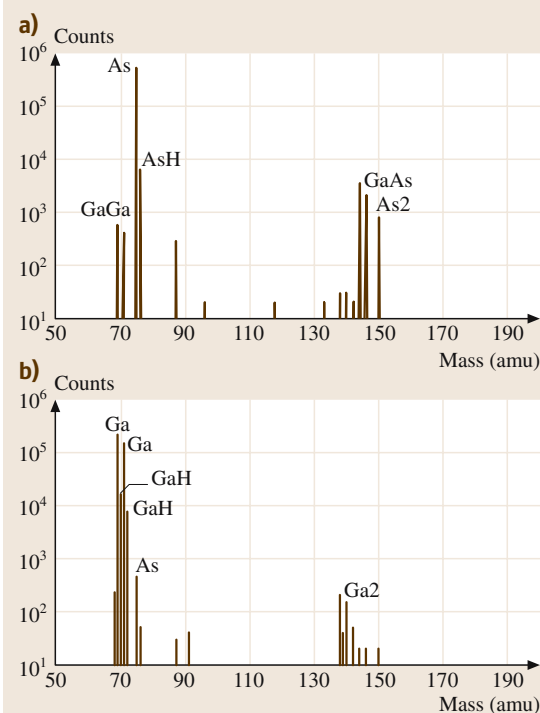


Fig. 18.9a,b Mass spectra from a GaAs wafer recorded (a) using Cs^+ primary ion bombardment and negative secondary ion detection, and (b) using O_2^+ primary ion bombardment and positive secondary ion detection

ment of very low energy primary ion beam columns to enable the characterization of shallow implants.

Clearly, as is illustrated in Fig. 18.8, the number of positive ions produced in the sputtering process depends upon the concentration of oxygen in the sample surface. In order to maximize the positive secondary ion yield, the sample surface needs to be saturated with oxygen, and this can be achieved with normal-incidence primary ion beams or by deliberately allowing oxygen to flood the sample surface as well as using an oxygen primary ion beam. In the example shown in Fig. 18.8, the primary ion beam was incident at approximately 45° and so complete oxidation of the silicon surface was not achieved, meaning that the silicon and oxygen signals in the equilibrium region are not as high as from the fully oxidized native oxide layer.

Not all species prefer to form positive ions in the sputtering process; some prefer to produce negative ions. Oxygen is a case in point, and indeed it would be difficult to conduct an analysis for oxygen using an oxygen beam. For electronegative species, it turns out that high

Fig. 18.10a,b Reference sample of boron ion-implanted silicon, (a) raw data profile and (b) quantified depth profile

negative secondary ion yields can be achieved if the surface is sputtered using cesium ions. An illustration of the relative ion yields of gallium and arsenic as positive and negative ions from a gallium arsenide surface with oxygen and cesium ion bombardment are shown in Fig. 18.9. Some species can be reluctant to produce either positive or negative secondary ions, but in some cases (such as zinc), good sensitivity can be achieved by using cesium ion bombardment and monitoring the cesium–element molecular ion, CsZn^+ in the case of zinc.

Despite the wide variations in secondary ion yields from element to element and matrix to matrix, the quantification of impurities in semiconductors can be relatively straightforward when reference materials are used. For example, Fig. 18.10a shows a raw data depth profile of ^{11}B ions implanted in silicon; knowing the areal dose of ions implanted into the sample and the depth of the crater sputtered into the sample, it is a relatively simple task (usually buried in the instrument software) to convert boron counts into concentrations

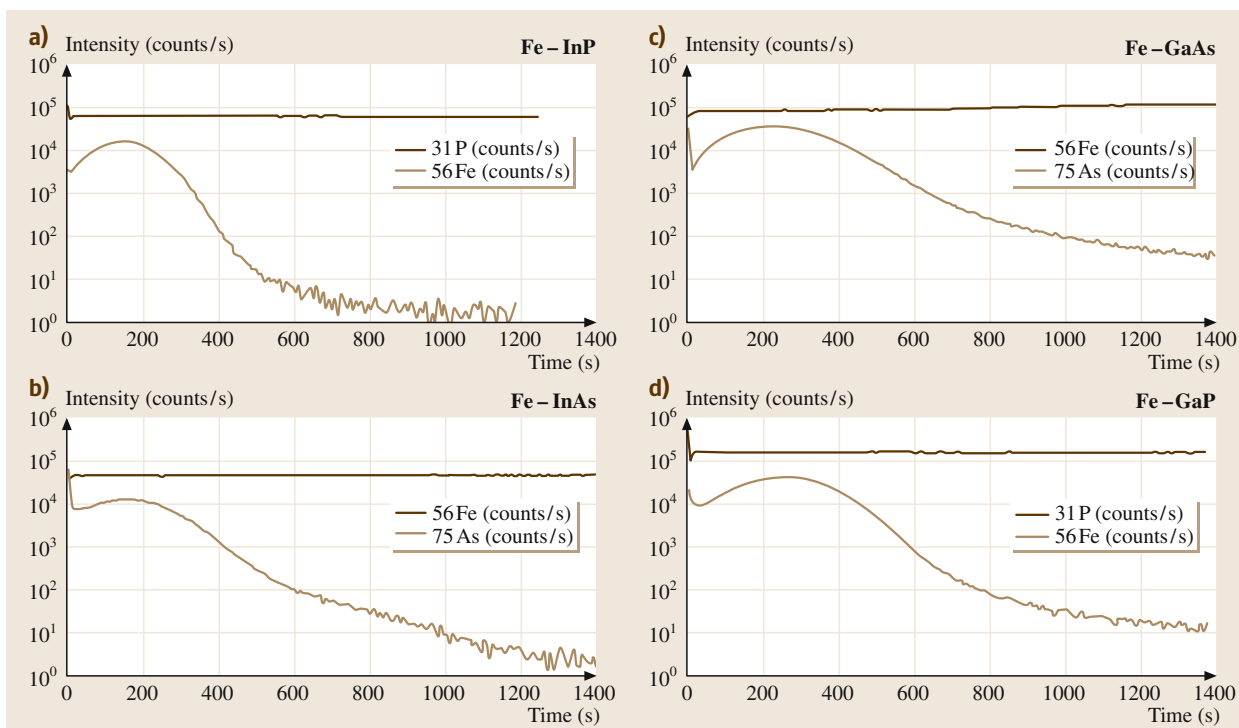
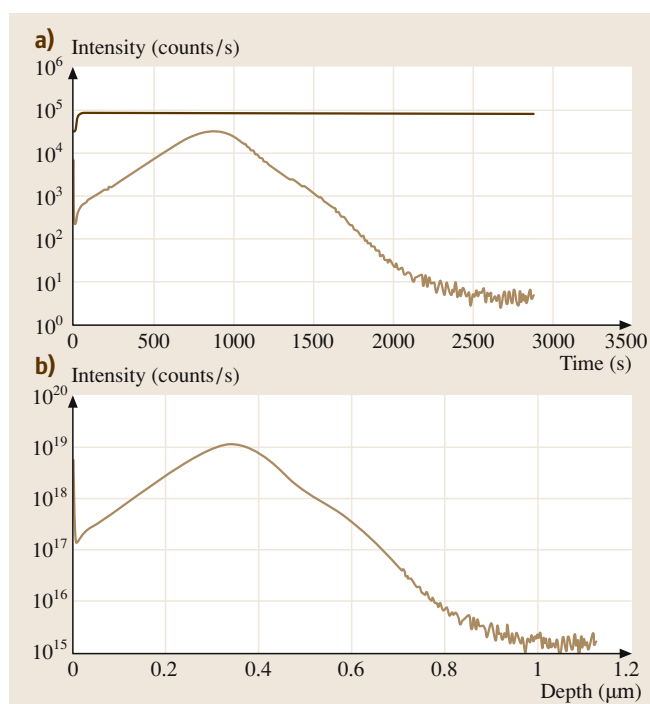


Fig. 18.11 (a)–(d) show iron profiles, recorded under identical conditions, from implants into InP, InAs, GaAs and GaP, respectively

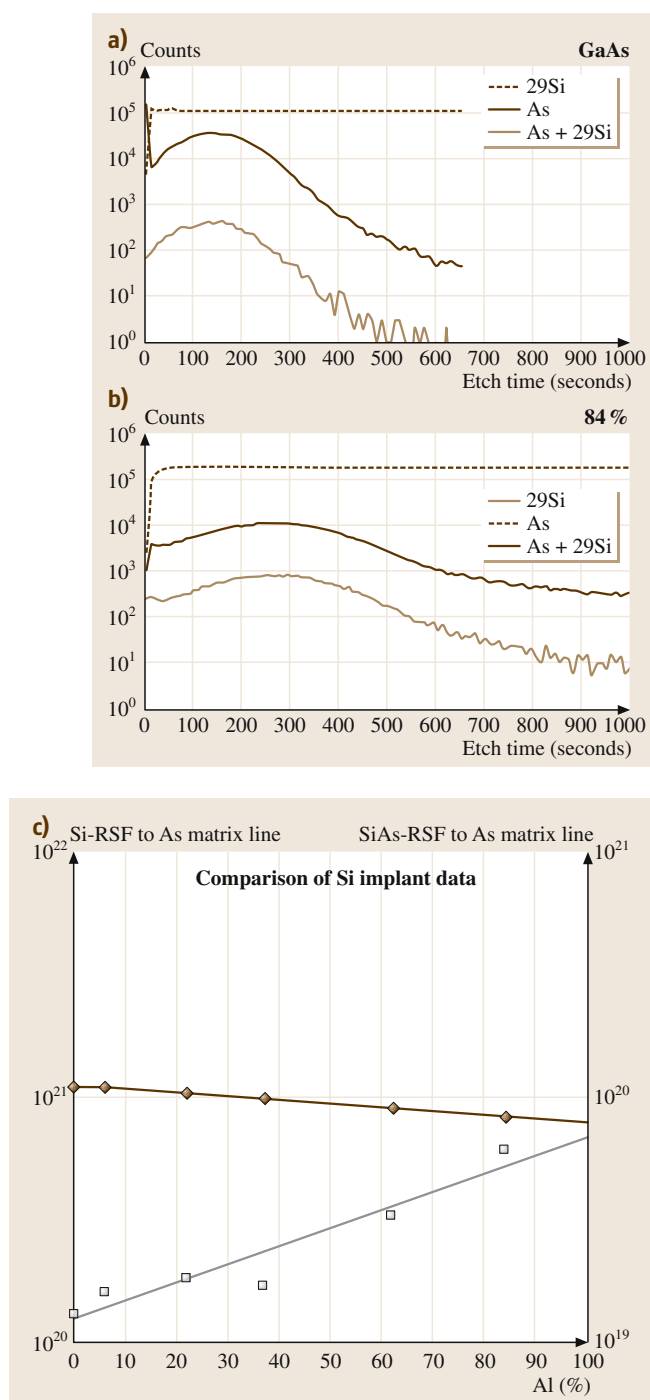


Fig. 18.12a–c Depth profiles of silicon as Si⁻ and AsSi⁻ in (a) GaAs and (b) Ga_{0.16}Al_{0.84}As, and (c) variation in relative sensitivity factor as a function of aluminium content

(Fig. 18.10b). By recording a signal from the matrix element, either during the profile as in Fig. 18.10a, or in the crater after the profile, a relative sensitivity factor for the analyte (boron) in the matrix of interest (silicon) can be obtained which can then be used to determine the concentration of that analyte in another sample of the same matrix. Errors or uncertainties can arise from both the crater depth measurement and the implanted dose measurement. At the time of writing, there are only three metrologically traceable reference samples for SIMS (boron, arsenic and phosphorus in silicon) produced by NIST in the USA. For other species and other matrices, the analyst must rely upon locally produced materials with no traceability.

Care must be exercised when the analyte of interest is present in different matrices. For example, Figs. 18.11a–c and d show iron profiles, recorded under identical conditions (positive ion detection with O₂⁺ primary ion bombardment), from implants into InP, InAs, GaAs and GaP produced in the same implant run. While the relative sensitivity of iron to the phosphorus matrix signal is the same for InP and GaP matrices, there is a variation of the order of 20% in the arsenide matrices. The useful ion yield of iron (the total number of ions produced as a fraction of the number of atoms present) increases approximately four-fold in the gallium-based matrices compared to the indium-based matrices. The profile in the GaAs sample, Fig. 18.11c, also shows a further complication: the arsenic matrix signal increases at about 1000 s into the profile. This could be interpreted as an upwards drift in the primary ion beam current, but it is actually caused by roughening of the GaAs surface as a result of the sputtering process. This effect is more common with oxygen primary ion bombardment than with cesium bombardment, and it has been shown that it is possible to reduce such effects by rotating the sample during analysis. Not only does the roughening modify the analytical signal, it also causes an apparent broadening of sharp features present in the sample as the surface roughness is convoluted with the feature width. The problems associated with sputter-induced roughness can become more severe when dealing with metal films but, once again, can be overcome by sample rotation.

Another example of how relative sensitivity factors vary with the matrix is shown in Fig. 18.12. Here a comparison is shown between the behavior of the Si⁻ and AsSi⁻ signals as the matrix is changed from GaAs to Ga_{0.16}Al_{0.84}As. Molecular signals such as AsSi⁻ are sometimes used to give increased signal compared to the atomic signal but, as in the example shown, the backgrounds are often similarly increased. What is inter-

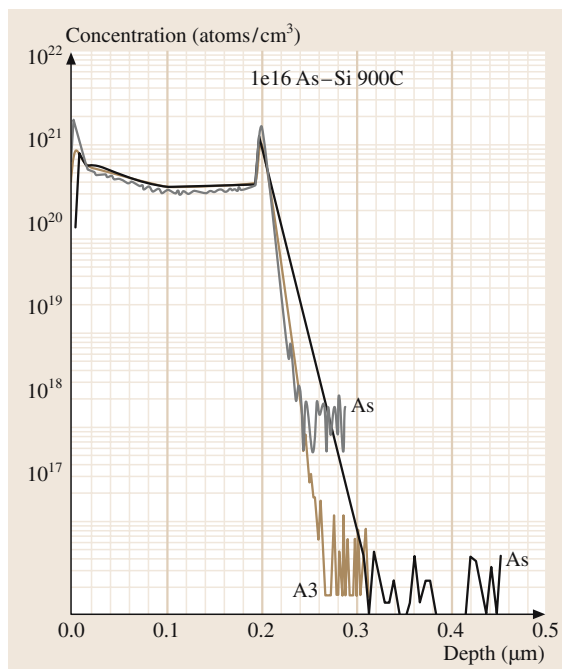


Fig. 18.13 Depth profiles of As in Si for different analytical conditions

esting here is that whilst the ion yields from the Si^- and As^- signals increase, but at different rates with increasing Al content, the AsSi^- signals decrease. Ideally the analyst needs to be aware of, and able to take account of, such variations in response. Simply using the relative sensitivity factor derived for a GaAs matrix and applying this to high Al content material using a point-by-point normalization to the As^- signal can lead to a 30% overestimate of the silicon content. However, monitoring the AsSi^- signal instead and using the same normalization procedure can underestimate the silicon content by over 350%.

With an ideal analytical technique the analytical signal will vary linearly with concentration over all concentrations. With SIMS, for uniform matrices, the measured signal is indeed linear with concentration for dilute systems. However, once the analyte concentration begins to exceed a few percent of the material, the signal may begin to vary in a nonlinear way with concentration. Thus, whilst SIMS is the ideal technique for

quantifying minor and trace contaminants, it is less than ideal for determining matrix element concentrations. The most successful method of measuring elemental concentrations at high levels was achieved by monitoring the cesium molecular species (CsAl^+ and CsGa^+ and CsAs^+) in GaAlAs. In GaAlAs the ($\text{CsAl}^+/\text{CsAs}^+$) and ($\text{CsGa}^+/\text{CsAs}^+$) ratios do vary linearly with the Al content.

Returning to practical analysis, another factor that needs to be considered in relation to DSIMS, and one that is related to the primary bombarding particles, is atomic mixing. The primary ions used to bombard the surface transfer momentum to the atoms in the sample surface. While some of the atoms will be sputtered from the surface, others will be driven deeper into the sample, thus broadening any concentration distributions that were present within the original sample. The higher the primary ion beam energy, the more severe the atomic mixing. Consequently, as the need to analyze shallower and shallower structures has increased, there has been a move to develop instruments that use lower bombardment energies and sources producing polyatomic ions. A simple example of how the analytical conditions can cause profile distortion is shown in Fig. 18.13. These profiles were recorded from an arsenic-implanted and annealed polysilicon layer on a thin oxide on silicon. Using an oxygen primary ion beam striking the surface at oblique incidence with an impact energy of 5.5 keV (a) produces a profile that shows a spike followed by a rapid fall at the polysilicon/oxide interface. Using a cesium primary ion beam striking the surface at close to normal incidence with an impact energy of 14.5 keV (b) gives a much better detection limit with smaller interface spikes but with a much less abrupt fall in the arsenic signal. Profile (c), which uses the same bombardment conditions as profile (a) but combined with oxygen flooding of the sample surface (and consequently took longer to acquire), shows good agreement with profile (b) in terms of the size of the interface spike – which is exaggerated in profile (a) because of the increased ion yield from the oxide layer compared to the polysilicon – and good agreement with profile (a) in terms of the decay of the arsenic signal in the interface region, and has better sensitivity than profile (a) but less than profile (b). These profiles demonstrate the trade-off that the analyst must make when choosing analysis conditions.

18.4 Conclusion

The various surface chemical analysis techniques have their own strengths and weaknesses. No one method is suitable for all of the tasks the analyst faces; sometimes one technique is sufficient to address the problem

at hand, sometimes a combination of them is required. However, the approach should be successful if the technique(s) is (are) fit for the purpose of the task.

19. Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications

The chapter provides a summary of the fundamental concepts that are needed to understand the heat capacity C_p , thermal conductivity κ , and thermal expansion coefficient α_L of materials. The C_p , κ , and α of various classes of materials, namely, semiconductors, polymers, and glasses, are reviewed, and various typical characteristics are summarized. A key concept in crystalline solids is the Debye theory of the heat capacity, which has been widely used for many decades for calculating the C_p of crystals. The thermal properties are interrelated through Grüneisen's theorem. Various useful empirical rules for calculating C_p and κ have been used, some of which are summarized. Conventional differential scanning calorimetry (DSC) is a powerful and convenient thermal analysis technique that allows various important physical and chemical transformations, such as the glass transition, crystallization, oxidation, melting etc. to be studied. DSC can also be used to obtain information on the kinetics of the transformations, and some of these thermal analysis techniques are summarized. Temperature-modulated DSC, TMDSC, is a relatively recent innovation in which the sample temperature is ramped slowly and, at the same time, sinusoidally modulated. TMDSC has a number of distinct advantages compared with the conventional DSC since it measures the complex heat capacity. For example, the glass-transition temperature T_g measured by TMDSC has almost no dependence on the thermal history, and corresponds to an almost step life change in C_p .

19.1 Heat Capacity	386
19.1.1 Fundamental Debye Heat Capacity of Crystals	386
19.1.2 Specific Heat Capacity of Selected Groups of Materials ...	388
19.2 Thermal Conductivity	391
19.2.1 Definition and Typical Values	391
19.2.2 Thermal Conductivity of Crystalline Insulators	391
19.2.3 Thermal Conductivity of Noncrystalline Insulators	393
19.2.4 Thermal Conductivity of Metals ...	395
19.3 Thermal Expansion	396
19.3.1 Grüneisen's Law and Anharmonicity	396
19.3.2 Thermal Expansion Coefficient α ..	398
19.4 Enthalpic Thermal Properties	398
19.4.1 Enthalpy, Heat Capacity and Physical Transformations	398
19.4.2 Conventional Differential Scanning Calorimetry (DSC)	400
19.5 Temperature-Modulated DSC (TMDSC)	403
19.5.1 TMDSC Principles	403
19.5.2 TMDSC Applications	404
19.5.3 Tzero Technology	405
References	406

The new Tzero DSC has an additional thermocouple to calibrate better for thermal lags inherent in the DSC measurement, and allows more accurate thermal analysis.

The selection and use of electronic materials, one way or another, invariably involves considering such thermal properties as the specific heat capacity (c_s), thermal conductivity (κ), and various thermodynamic and structural transition temperatures, for example, the melting or fusion temperature (T_m) of a crystal, glass transformation (T_g) and crystallization temperature (T_c) for glasses and amorphous polymers. The thermal expansion coefficient (α) is yet another important material property that comes into full play in applications of electronic mater-

ials inasmuch as the thermal expansion mismatch is one of the main causes of electronic device failure. One of the most important thermal characterization tools is the differential scanning calorimeter (DSC), which enables the heat capacity, and various structural transition temperatures to be determined. Modulated-temperature DSC in which the sample temperature is modulated sinusoidally while being slowly ramped is a recent powerful thermal analysis technique that allows better thermal characterization and heat-capacity measurement. In addition, it

can be used to measure the thermal conductivity. The present review is a selected overview of thermal properties and the DSC technique, in particular MTDSC. The overview is written from a materials science perspective with emphasis on phenomenology rather than fundamental physics.

19.1 Heat Capacity

19.1.1 Fundamental Debye Heat Capacity of Crystals

The heat capacity of a solid represents the increase in the enthalpy of the crystal per unit increase in the temperature. The heat capacity is usually defined either at constant volume or at constant pressure, C_V and C_P , respectively. C_V represents the increase in the internal energy of the crystal when the temperature is raised because the heat added to the system increases the internal energy U without doing mechanical work by changing the volume. On the other hand C_P represents the increase in the enthalpy H of the system per unit increase in the temperature. Thus,

$$C_V = \left(\frac{\partial H}{\partial T} \right)_V = \left(\frac{\partial U}{\partial T} \right)_V \quad \text{and} \quad C_P = \left(\frac{\partial H}{\partial T} \right)_P. \quad (19.1)$$

The exact relationship between C_V and C_P is

$$C_V = C_P - \frac{T\alpha^2}{\rho K}, \quad (19.2)$$

where T is the temperature, ρ is the density, α is the linear expansion coefficient and K is the compressibility. For solids, C_V and C_P are approximately the same. The increase in the internal energy U is due to an increase in the energy of lattice vibrations. This is generally true for all solids except metals at very low temperatures where the heat capacity is due to the conduction electrons near the Fermi level becoming excited to higher energies. For most practical temperature ranges of interest, the heat capacity of most solids is determined by the excitation of lattice vibrations. The molar heat capacity C_m is the increase in the internal energy U_m of a crystal of Avogadro's number N_A atoms per unit increase in the temperature at constant volume, that is, $C_m = (dU_m/dT)_V$. The Debye heat capacity is still the most successful model for understanding the heat capacity of crystals, and is based on the thermal excitation of lattice vibrations, that is phonons, in the crystal [19.4];

The thermal properties of a large selection of materials can be found in various handbooks [19.1, 2]. In the case of semiconductors, *Adachi's* book is highly recommended [19.3] since it provides useful relationships between the thermal properties for various group IV, II–V and II–VI semiconductors.

it is widely described as a conventional heat capacity model in many textbooks [19.5, 6]. The vibrational mean energy at a frequency ω is given by

$$\bar{E}(\omega) = \frac{\hbar\omega}{\exp(\frac{\hbar\omega}{k_B T}) - 1}, \quad (19.3)$$

where k_B is the Boltzmann constant. The energy $\bar{E}(\omega)$ increases with temperature. Each phonon has an energy of $\hbar\omega$ so that the phonon concentration in the crystal increases with temperature. To find the internal energy due to all the lattice vibrations we must also consider how many vibrational modes there are at various frequencies. That is, the distribution of the modes over the possible frequencies: the spectrum of the vibrations. Suppose that $g(\omega)$ is the number of modes per unit frequency, that is, $g(\omega)$ is the vibrational density of states or modes. Then $g(\omega)d\omega$ is the number of vibrational states in the range $d\omega$. The internal energy U_m of all lattice vibrations for 1 mole of solid is,

$$U_m = \int_0^{\omega_{\max}} \bar{E}(\omega)g(\omega)d\omega. \quad (19.4)$$

The integration is up to certain allowed maximum frequency ω_{\max} . The density of states $g(\omega)$ for the lattice vibrations in a periodic three-dimensional lattice, in a highly simplified form, is given by

$$g(\omega) \approx \frac{3}{2\pi^2} \frac{\omega^2}{v^3}, \quad (19.5)$$

where v is the mean velocity of longitudinal and transverse waves in the solid. The maximum frequency is ω_{\max} and is determined by the fact that the total number of modes up to ω_{\max} must be $3N_A$. It is called the *Debye frequency*. Thus, integrating $g(\omega)$ up to ω_{\max} we find,

$$\omega_{\max} \approx v(6\pi^2 N_A)^{1/3}. \quad (19.6)$$

This maximum frequency ω_{\max} corresponds to an energy $\hbar\omega_{\max}$ and to a temperature T_D defined by,

$$T_D = \frac{\hbar\omega_{\max}}{k}, \quad (19.7)$$

and is called the *Debye temperature*. Qualitatively, it represents the temperature above which all vibrational frequencies are executed by the lattice waves.

Thus, by using (19.3) and (19.5) in (19.4) we can evaluate U_m and hence differentiate U_m with respect to temperature to obtain the molar heat capacity at constant volume,

$$C_m = 9R \left(\frac{T}{T_D} \right)^3 \int_0^{T_D/T} \frac{x^4 e^x dx}{(e^x - 1)}, \quad (19.8)$$

which is the well-known Debye heat capacity expression.

Figure 19.1 represents the constant-volume Debye molar heat capacity C_m for a perfect crystal (19.8) as a function of temperature, normalized with respect to the Debye temperature. The well-known classical *Dulong–Petit rule* ($C_m = 3R$) is only obeyed when $T > T_D$. The Dulong–Petit (DP) rule is a direct consequence of the applications of Maxwell’s theorem of equipartition of energy and the classical kinetic molecular theory to vibrations of atoms in a crystal. Notice that C_m at $T = 0.5T_D$ is $0.825(3R)$, whereas at $T = T_D$ it is $0.952(3R)$. For most practical purposes, C_m is within 6% of $3R$ when the temperature is $0.9T_D$. For example, for copper $T_D = 315$ K and above about $0.9T_D$, that is above 283 K (or 10°C), $C_m \approx 3R$, as borne out by experiments. Table 19.1 provides typical values for T_D , and heat capacities for a few selected elements. At the lowest temperatures, when $T \ll T_D$, (19.8) predicts that $C_m \propto T^3$ and this is indeed observed in low-temperature heat capacity experiments on a variety of crystals. On the other hand, well-known exceptions are glasses, noncrystalline solids, whose heat capacity is proportional to $a_1T + a_2T^3$, where a_1 and a_2 are constants (e.g. [19.6, 7]).

It is useful to provide a physical picture of the Debye model inherent in (19.8). As the temperature increases from near zero, the increase in the crystal’s vibrational energy is due to more phonons being created and higher frequencies being excited. The phonon concentration increases as T^3 and the mean phonon energy increases as T . Thus, the internal energy increases as T^4 . At temperatures above T_D , increasing the temperature creates more phonons but does not increase the mean phonon energy and does not excite higher frequencies. All frequencies up to ω_{\max} have been now excited. The internal

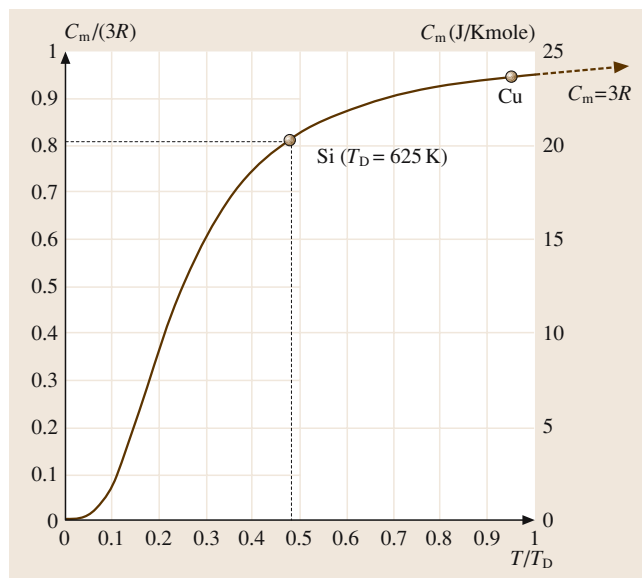


Fig. 19.1 Debye constant-volume molar heat capacity curve. The dependence of the molar heat capacity C_m on temperature with respect to the Debye temperature: C_m versus T/T_D . For Si, $T_D = 625$ K, so at room temperature (300 K), $T/T_D = 0.48$ and C_m is only $0.81(3R)$

energy increases only due to more phonons being created. The phonon concentration and hence the internal energy increases as T ; the heat capacity is constant, as expected from (19.8).

In general terms, elements with higher atomic mass tend to have lower T_D , and their molar heat capacities are closer to the DP limit. Usually T_D is taken to depend on the temperature to accurately model the experimental C_p – T behavior; the dependence of T_D on T is only important at low temperatures, and can be neglected in most practical applications in engineering.

There are two important assumptions in the Debye theory. First is that all thermal excitations are lattice vibrations (i.e., involving phonons), which is usually the case in solids. Secondly, the interatomic potential-energy (PE) curve as a function of interatomic displacement x from equilibrium, is assumed to be parabolic, that is *harmonic* or symmetrical: $\text{PE} = \text{PE}_{\min} + (1/2)\beta x^2$, where β is a constant (the spring constant in the equation of net force versus displacement, $F = -\beta x$). Within the harmonic approximation, the lattice vibrations are *plane waves* that are independent of each other; they are vibrational normal modes of the lattice. Further, a symmetrical PE curve does not allow thermal expansion, which means that C_V

Table 19.1 Debye temperatures (T_D), heat capacities, thermal conductivities and linear expansion coefficients of various selected metals and semiconductors. C_m , c_s , κ , and α are at 25 °C. For metals, T_D is obtained by fitting the Debye curve to the experimental molar heat capacity data at the point $C_m = \frac{1}{2}(3R)$. T_D data for metals from [19.8]. Other data from various references, including [19.2] and the Goodfellow metals website

Metals	Ag	Al	Au	Bi	Cu	Ga	Hg	In	Pd	W	Zn
T_D (K)	215	394	170	120	315	240	100	129	275	310	234
C_m (J/K mol)	25.6	24.36	25.41	25.5	24.5	25.8	27.68	26.8	25.97	24.45	25.44
c_s (J/K g)	0.237	0.903	0.129	0.122	0.385	0.370	0.138	0.233	0.244	0.133	0.389
κ (W/m K)	420	237	317	7.9	400	40.6	8.65	81.6	71.8	173	116
α (K ⁻¹) $\times 10^{-6}$	19.1	23.5	14.1	13.4	17	18.3	61	24.8	11	4.5	31
Semiconductors	Diamond	Si	Ge	AlAs	CdSe	GaAs	GaP	InAs	InP	ZnSe	ZnTe
T_D (K)	1860	643	360	450	135	370	560	280	425	340	260
C_m (J/K mol)	6.20	20.03	23.38	43.21	53.77	47.3	31.52	66.79	46.95	51.97	49.79
c_s (J/K g)	0.540	0.713	0.322	0.424	0.281	0.327	0.313	0.352	0.322	0.360	0.258
κ (W/m K)	1000	156	60	91	4	45	77	30	68	19	18
α (K ⁻¹) $\times 10^{-6}$	1.05	2.62	5.75	4.28	7.43	6.03	4.89	5	4.56	7.8	8.33

and C_p are identical. However, the actual interatomic PE is anharmonic, that is, it has an additional x^3 term. It is not difficult to show that in this case the vibrations or phonons interact. For example, two phonons can *mix* to generate a third phonon of higher frequency or a phonon can decay into two phonons of lower frequency etc. Further, the anharmonicity also leads to thermal expansion, so that C_V and C_p are not identical as is the case in the Debye model. As a result of the anharmonic effects, C_p continues to increase with temperature beyond the $3R$ Dulong–Petit rule, though the increase with temperature is usually small.

19.1.2 Specific Heat Capacity of Selected Groups of Materials

Many researchers prefer to quote the heat capacity for one mole of the substance, that is quote C_m , and sometime express C_m in terms of R . The limit $C_m = 3R$ is the DP rule. It is not unusual to find materials for which C_m can exceed the $3R$ limit at sufficiently high temperatures for a number of reasons, as discussed, for example, by Elliott [19.6]. Most applications of electronic materials require a knowledge of the specific heat capacity c_s , the heat capacity per unit mass. The heat capacity per unit volume is simply c_s/ρ , where ρ is the density. For a crystal that has only one type of atom with an atomic mass M_{at} (g/mol) in its unit cell (e.g. Si), c_s is C_m/M_{at} expressed in J/K g.

While the Debye heat capacity is useful in predicting the molar heat capacity of a crystal at any temperature, there are many substances, such as metals, both pure

metals and alloys, and various semiconductors (e.g. Ge, CdSe, ZnSe etc.) and ionic crystals (e.g. CsI), whose room-temperature heat capacities approximately follow the simple DP rule of $C_m = 3R$, the limiting value in Fig. 19.1. For a metal alloy, or a compound such as $A_xB_yC_z$, that is made up of three components A , B and C with molar fractions x , y and z , where $x + y + z = 1$, the overall molar heat capacity can be found by adding individual molar heat capacities weighted by the molar fraction of the component,

$$C_m = xC_{mA} + yC_{mB} + zC_{mC}, \quad (19.9)$$

where C_{mA} , C_{mB} and C_{mC} are the individual molar heat capacities. Equation (19.9) is the *additive rule of molar heat capacities*. The corresponding specific heat capacity is

$$c_s = 3R/\bar{M}_{at}, \quad (19.10)$$

where $\bar{M}_{at} = xM_A + yM_B + zM_C$ is the mean atomic mass of the compound, and M_A , M_B and M_C are the atomic masses of A , B and C . For example, for ZnSe, the average mass $\bar{M}_{at} = (1/2)(78.96 + 65.41) = 72.19$ g/mol, and the DP rule predicts $c_s = 3R/\bar{M}_{at} = 0.346$ J/Kg, which is almost identical to the experimental value at 300 K.

The modern Debye theory and the classical DP rule that $C_m = 3R$ are both based on the addition of heat increasing the vibrational energy of the atoms or molecules in the solid. If the molecules are able to rotate, as in certain polymers and liquids, then the molar heat capacity will be more than $3R$. For example, the heat capacity of

molten Bi–BiI₂ mixture is more than the DP rule, and the difference is due to the rotational energies associated with various molecular units in the molten state [19.9].

The heat capacity of polymers are usually quite different than the expected simple Debye heat capacity behavior. Most polymers are not fully crystalline but either amorphous or semicrystalline, a mixture of amorphous and crystalline regions. Secondly, and most significantly, the polymer structures usually have a main chain (a backbone) and various side groups attached to this backbone. Stretching and possible bending vibrations and possible partial rotations of these side groups provide additional contributions to the overall heat capacity. The extent of freedom the main chain and side groups have in executing stretching and bending vibrations and rotations depends on whether the structure is above or below the glass-transition temperature T_g . (The glass-transition temperature is defined and discussed later.) Above T_g , the polymer structure is floppy and has sufficient free volume for various molecular motions to be able to absorb the added heat. The heat capacity is therefore larger above T_g than that below T_g , where the structure is frozen.

The heat capacity of most polymers increases with temperature. For most solid polymers (below the glass-transition temperature), the fractional rate of increase in the heat capacity with temperature follows

$$(1/C_P)(dC_P/dT) \approx 3 \times 10^{-3} \text{ K}^{-1} \quad (T < T_g) \quad (19.11)$$

to within about 5%. This empirical rule is particularly useful for estimating C_P at a desired temperature from a given value at another temperature e.g. at room temperature. For liquid polymers (above T_g) on the other hand, $(1/C_P)(dC_P/dT)$ is less ($\approx 1 \times 10^{-3} \text{ K}^{-1}$) and has a large variability (more than 30%) from polymer to polymer. There are various empirical rules for predicting the heat capacity of a polymer based on the backbone chain structure and types of side groups, as discussed by *van Krelen* and *Hoflyzer* [19.10]. The most popular is to simply sum heat capacity contributions per mole from each distinguishable molecular unit in the polymer structure's repeat unit. The heat capacity contribution from each type of identifiable molecular unit (e.g. $-\text{CH}_2-$ in the backbone or $-\text{CH}_3$ side group) in polymer structures are listed in various tables. For example, the molar heat capacity of polypropylene can be calculated approximately by adding contributions from three distinguishable units, CH_2 , CH and CH_3 in the repeating unit that makes up the polypropylene structure, as illustrated schematically in Fig. 19.2. Once the

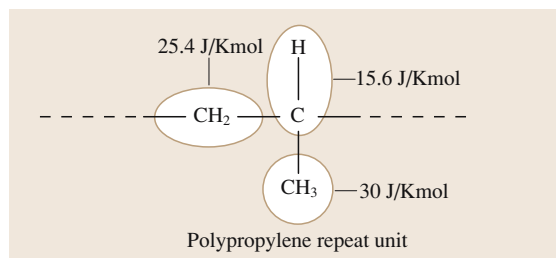


Fig. 19.2 Molar heat capacity of solid polypropylene calculated from various distinguishable units in the repeat unit

contributions are added, the sum is 71.9 J/K mol. The molecular mass of the repeat unit corresponds to three C atoms and six H atoms, or to 42 g/mol so that the specific heat is 71.9 J/K mol/42 g/mol or 1.71 J/Kg, which is close to the experimental value. (Polypropylene is normally semicrystalline and its specific heat also depends on the crystalline-to-amorphous phase ratio.) The term *heat capacity per mole* in polymers normally refers to a mole of repeat units or segments in the polymer.

More rigorous characterization of the heat capacity of polymers tends to be complicated by the fact that most polymers are a mixture of the crystalline and amorphous phases, which have different C_P versus T characteristics, and the exact volume fractions of the phases may not be readily available. There have been a number of examples recently where the heat capacity characteristics of semicrystalline polymers have been critically examined [19.11]. A comprehensive overview of the heat capacity of polymers has been given by *Wunderlich* [19.12], which is highly recommended.

The overall or effective heat capacity of polymer blends can be usually calculated from the individual capacities and the volume fractions of the constituents. The simplest is a linear mixture rule in which individual capacities are weighted by the volume fraction and then summed.

The heat capacity of glasses depends on the composition and increases with temperature. There is a sharp change in the dependence of C_P on T through the glass-transition temperature T_g . For example, for zirconium, barium, lanthanum, aluminum, sodium (ZBLAN) glasses, C_m is 24 J/K mol below T_g and slightly less than $3R$. On the other hand, C_m is 38.8 J/K mol above T_g , and greater than $3R$. (The C_P change at T_g is discussed below.) The dependence of C_P on T below T_g has been modeled by a number of expressions dating back to the 1950s. For example, *Sharp* and *Ginther* [19.13] in 1951 proposed the following empirical expressions

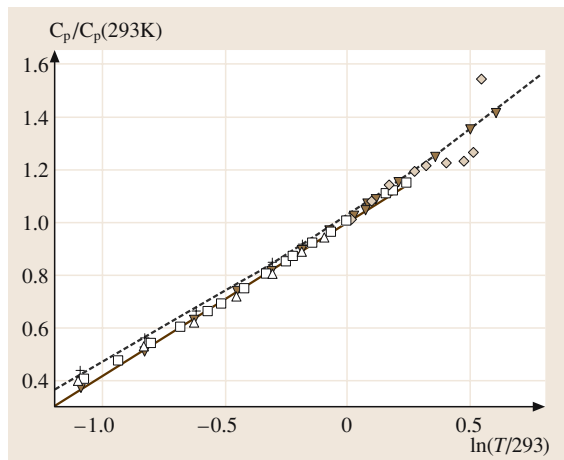


Fig. 19.3 Normalized heat capacity $C_p(T)/C_p(298)$ versus $\ln(T/298)$ for various glasses. *Open square*: 12.5Na₂O, 12.5Al₂O₃, 75SiO₂; *solid circle*: 25MgO, 25CaO, 50SiO₂; *open triangle*: 33Na₂O, 67B₂O₃; *solid triangle*: 24BaO, 1Fe₂O₃, 75B₂O₃; *open diamond*: B₂O₃ (includes T_g); +30K₂O, 70GeO₂. Adapted from *Khalimovskaya-Churkina and Priven* [19.14]. The *solid straight line* is (19.14), whereas the *dashed line* is the experimental trend line

for the mean and true specific heat capacity c_m and c respectively, of oxide glasses in the range 0–1300 °C,

$$c_m = \frac{H_2 - H_1}{T_2 - T_1} = \frac{aT + c_o}{0.00146T + 1} \quad \text{and} \quad (19.12a)$$

$$c = \frac{dH}{dT} = \frac{aT + c_m}{0.00146T + 1} \quad (19.12b)$$

in which T is in °C, and a and c_o are calculated from the glass composition as described in the original paper [19.13]. The mean heat capacity in (19.12)a then predicts the heat capacity to within about 1%. More recently, for various oxide glasses from 300 K to T_g , *Inaba et al.* [19.16] have proposed an exponential empirical expression of the form,

$$C_p = 3R \left[1 - \exp \left(-1.5 \frac{T}{T_D} \right) \right], \quad (19.13)$$

where T_D is the Debye temperature. Equation (19.13) predicts a lower C_p than the Debye equation itself at the same temperature for the same T_D . There are also other empirical expressions for C_p (for example, a three-term polynomial in T) that typically involve at least three fitting parameters. *Khalimovskaya-Churkina* and

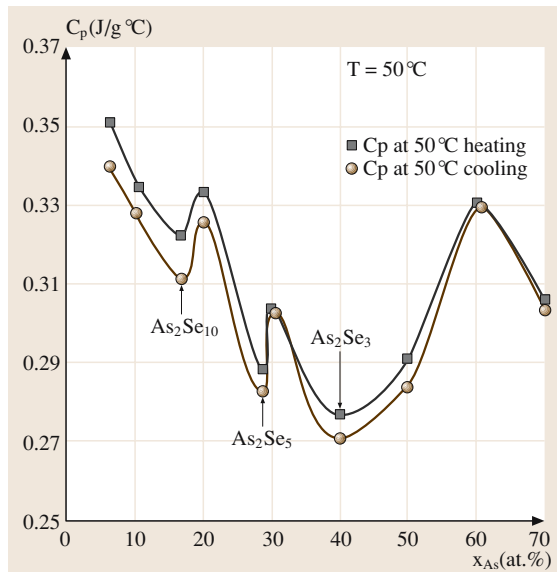


Fig. 19.4 The specific heat capacity versus As content in As_xSe_{1-x} . (After *Wagner and Kasap* [19.15])

Priven [19.14] have shown that a simple two-parameter empirical expression of the form

$$C_p = a + b \ln T, \quad (19.14)$$

where a and b are composition-dependent parameters, can be fruitfully used to represent the dependence of C_p on T from 100 K to the lower boundary of T_g for various oxide glasses to within about 2.5% error as shown in Fig. 19.3. The authors further show that, to within about 4%,

$$C_p = C_p(293 \text{ K}) [1 + 0.52 \ln(T/293)] . \quad (19.15)$$

What is significant in this equation is that the composition dependence of C_p has been incorporated into $C_p(293 \text{ K})$, and once the latter is calculated, the temperature dependence of C_p is then given by (19.15).

Various empirical rules have been proposed for approximately calculating the heat capacity of oxide glasses from its components. The simplest is the additive rule stated in (19.9), in which the overall heat capacity per mole is found from adding individual molar heat capacities of the constituents, so-called partial molar heat capacities, weighted by their molar fractions. However, for a heterogeneous mixture, that is for a solid in which there are two or more distinct phases present, the specific heat capacity may involve more complicated calculations.

Figure 19.4 shows the heat capacity of an $\text{As}_x\text{Se}_{1-x}$ glass as the composition is varied. The specific heat capacity changes with composition and shows special features at certain critical compositions that correspond to the appearance of various characteristic molecular units in the glass, or the structure becoming optimally connected (when the mean coordination number $\langle r \rangle = 2.4$). With low concentrations of As, the structure is Se-rich and has a floppy structure, and the heat capacity

per mole is about $3R$. As the As concentration is increased, the structure becomes more rigid, and the heat capacity decreases, and eventually at 40 at % As, corresponding to As_2Se_3 , the structure has an optimum connectivity ($\langle r \rangle = 2.4$) and the heat capacity is minimum and, in this case, close to $2.51R$. There appears to be two minima, which probably correspond to $\text{As}_2\text{Se}_{10}$ and As_2Se_5 , that is to AsSe_5 and $\text{AsSe}_{5/2}$ units within the structure.

19.2 Thermal Conductivity

19.2.1 Definition and Typical Values

Heat conduction in materials is generally described by *Fourier's heat conduction law*. Suppose that J_x is the heat flux in the x -direction, defined as the quantity of heat flowing in the x -direction per unit area per unit second: the thermal energy flux. Fourier's law states that the heat flux at a point in a solid is proportional to the temperature gradient at that point and the proportionality constant depends on the material,

$$J_x = -\kappa \frac{dT}{dx}, \quad (19.16)$$

where κ is a constant that depends on the material, called the thermal conductivity (W/m K or $\text{W/m}^\circ\text{C}$), and dT/dx is the temperature gradient. Equation (19.16) is called Fourier's law, and effectively defines the thermal conductivity of a medium. Table 19.2 provides an overview of typical values for the thermal conductivity of various classes of materials.

The thermal conductivity depends on how the atoms in the solid transfer the energy from the hot region to the cold region. In metals, the energy transfer involves the conduction electrons. In nonmetals, the energy transfer involves lattice vibrations, that is atomic vibrations of the crystal, which are described in terms of phonons.

The thermal conductivity, in general, depends on the temperature. Different classes of materials exhibit different κ values and dependence of κ on T , but we can generalize very roughly as follows:

Most pure metals: $\kappa \approx 50\text{--}400 \text{ W/m K}$. At sufficiently high T , e.g. above $\approx 100 \text{ K}$ for copper, $\kappa \approx \text{constant}$. In magnetic materials such as iron and nickel, κ decreases with T .

Most metal alloys: κ lower than for pure metals; $\kappa \approx 10\text{--}100 \text{ W/m K}$. κ increases with increasing T .

Most ceramics: Large range of κ , typically $10\text{--}200 \text{ W/m K}$ with diamond and beryllia being exceptions with high κ . At high T , typically above $\approx 100 \text{ K}$, κ decreases with increasing T .

Most glasses: Small κ , typically less than $\approx 5 \text{ W/m K}$ and increases with increasing T . Typical examples are borosilicate glasses, window glass, soda-lime glasses, fused silica etc. Fused silica is noncrystalline SiO_2 with $\kappa \approx 2 \text{ W/m K}$.

Most polymers: κ is very small and typically less than 2 W/m K and increases with increasing T . Good thermal insulators.

19.2.2 Thermal Conductivity of Crystalline Insulators

In nonmetals heat transfer involves lattice vibrations, that is phonons. The heat absorbed in the hot region increases the amplitudes of the lattice vibrations which is the same as generating more phonons. These new phonons travel towards the cold regions and thereby transport the lattice energy from the hot to cold region. The thermal conductivity κ measures the rate at which heat can be transported through a medium per unit area per unit temperature gradient. It is proportional to the rate at which a medium can absorb energy, that is, κ is proportional to the heat capacity. κ is also proportional to the rate at which phonons are transported which is determined by their mean velocity v_{ph} . In addition, of

Table 19.2 Typical thermal conductivities of various classes of materials at 25 °C

Pure metals	Nb	Sn	Fe	Zn	W	Al	Cu	Ag
κ (W/m K)	53	64	80	116	173	237	400	420
Metal alloys	Stainless steel	55Cu–45Ni	Manganin (86Cu–12Mn–2Ni)	70Ni–30Cu	1080 Steel	Bronze (95Cu–5Sn)	Brass (63Cu–37Zn)	Dural (95Al–4Cu–1Mg)
κ (W/m K)	12–16	19.5	22	25	50	80	125	147
Ceramics and glasses	Glass-borosilicate	Silica-fused (SiO ₂)	S ₃ N ₄	Alumina (Al ₂ O ₃)	Magnesia (MgO)	Sapphire (Al ₂ O ₃)	Beryllia (BeO)	Diamond
κ (W/m K)	0.75	1.5	20	30	37	37	260	1000
Polymers	Poly-propylene	Polystyrene	PVC	Poly-carbonate	Nylon 6,6	Teflon	Polyethylene low density	Polyethylene high density
κ (W/m K)	0.12	0.13	0.17	0.22	0.24	0.25	0.3	0.5

course, κ is proportional to the mean free path Λ_{ph} that a phonon has to travel before losing its momentum, just as the electrical conductivity is proportional to the electron’s mean free path. A rigorous classical treatment gives κ as,

$$\kappa = \frac{1}{3} C_v v_{ph} \Lambda_{ph} , \tag{19.17}$$

where C_v is the heat capacity per unit volume. The mean phonon velocity v_{ph} is constant and approximately independent of temperature so that κ is governed primarily by C_v and Λ_{ph} . The mean free path Λ_{ph} depends on various processes that can scatter the phonons and hinder their propagation along the direction of heat flow. Phonons can collide with other phonons, crystal defects, impurities and crystal surfaces.

At high temperatures, κ is controlled by phonon–phonon scattering. The phonon–phonon collisions that are normally responsible for limiting the thermal conductivity, that is scattering the phonon momentum in the opposite direction to the heat flow, are due to the anharmonicity (asymmetry) of the interatomic potential

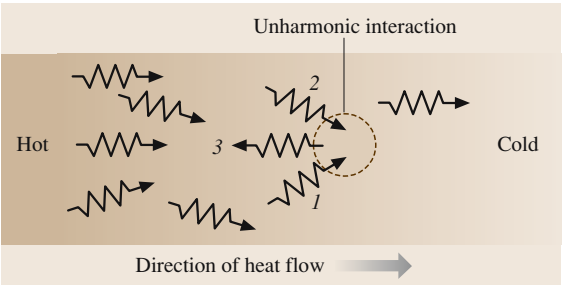


Fig. 19.5 Phonons generated in the hot region travel towards the cold region and thereby transport heat energy. Phonon–phonon anharmonic interaction generates a new phonon whose momentum is towards the hot region. (After [19.5])

energy curve. Stated differently, the net force F acting on an atom is not simply βx (where β is a spring constant) but also has an x^2 term, i.e., it is nonlinear. The greater the asymmetry or nonlinearity, the larger is the effect of such momentum-flipping collisions. The same asymmetry that is responsible for thermal expansion of solids is also responsible for determining the thermal conductivity. When two phonons 1 and 2 interact in a crystal region as in Fig. 19.5, the nonlinear behavior and the periodicity of the lattice, cause a new phonon 3 to be generated. This new phonon 3 has the same energy as the sum of 1 and 2 but it is traveling in the wrong direction. (The frequency of 3 is the sum of frequencies of 1 and 2.) The flipping of the phonon momentum is called an *Umklapp process* and is a result of the anharmonicity in the interatomic bond. Figure 19.6 shows the dependence of κ on T for three semiconductors. We can identify three types of behavior.

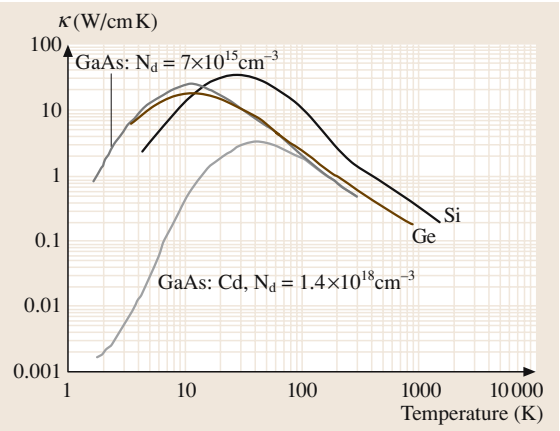


Fig. 19.6 Thermal conductivity of Si, Ge and GaAs. GaAs for two different levels of doping. Data extracted and replotted: Si [19.17]; Ge [19.18]; GaAs [19.19]

At low temperatures there are two factors that control κ . The phonon concentration is too low for phonon–phonon collisions to be significant. Instead, the mean free path Λ_{ph} is determined by phonon collisions with crystal imperfections, most significantly, crystal surfaces and grain boundaries. Thus, Λ_{ph} depends on the sample geometry and crystallinity. Further, as we expect from the Debye model, C_v depends on T^3 so that κ has the same temperature dependence as C_v , that is $\kappa \propto T^3$. Figure 19.6 shows the dependence of κ on T for Ge, Si and GaAs. Notice that κ increases with increasing T at low temperatures, and that κ also depends on the concentration of dopants (GaAs) or impurities in the crystal, which can scatter phonons. An extensive treatise on the thermal conductivity of semiconductors is given by Bhandari and Rowe [19.20]

As the temperature increases, the temperature dependence of κ becomes controlled by Λ_{ph} rather than C_v , which changes only slowly with T . The mean free path becomes limited by phonon–phonon collisions that obey the Umklapp process. These phonons have energies $\approx \frac{1}{2}k_B T_D$. The concentration of such phonons is therefore proportional to $\exp(T_D/2T)$. Since Λ_{ph} is inversely proportional to the Umklapp-obeying phonon concentration, κ decreases with increasing temperature, following a $\kappa \propto \exp(T_D/2T)$ type of behavior. At temperatures above the Debye temperature, C_v is, of course, constant, and the phonon concentration n_{ph} increases with temperature, $n_{\text{ph}} \propto T$. Thus, the mean free path decreases as $\Lambda_{\text{ph}} \propto 1/T$, which means that $\kappa \propto T^{-1}$, as observed for most crystals at sufficiently high temperatures. Figure 19.7 illustrates the expected dependence of κ on T for nonmetals, with three temperature regimes:

$$\kappa \propto \begin{cases} T^3; & \text{Low } T \\ \exp(-T_D/2T); & \text{Intermediate } T \\ T^{-1}; & \text{High } T. \end{cases} \quad (19.18)$$

Most crystalline semiconductors tend to follow this type of behavior. The thermal conductivity of important semiconductors has been discussed by Adachi [19.3]. In the temperature range above the peak of κ with T , Adachi used a $\kappa = AT^n$ type of power law to model the κ – T data, and has provided an extensive table for A and n for various group IV, III–V and II–V semiconductors [19.3]. For example, for GaAs, over $150 \leq T \leq 1500$, $A = 750$ and $n = -1.28$ so that $\kappa (\text{W/cm K}) = (750)T^{-1.28}$, where T is in Kelvin. Further, Adachi has been able to correlate κ with the lattice parameter a , mean atomic mass \bar{M} in the unit cell and the Debye temperature T_D .

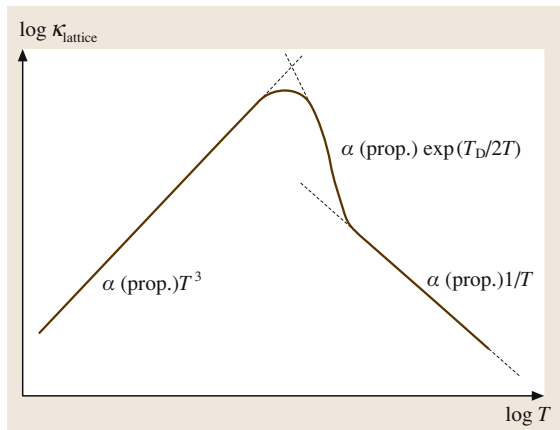


Fig. 19.7 Lattice thermal conductivity versus temperature (K) on a log–log plot. κ is due to phonons

19.2.3 Thermal Conductivity of Noncrystalline Insulators

Glasses

The thermal conductivity of glasses tends to be substantially smaller than their crystalline counterparts, and tends to increase with the temperature almost monotonically. The most striking difference at room temperature is that, as the temperature is decreased, κ for a glass decreases whereas for a crystal it increases. This dif-

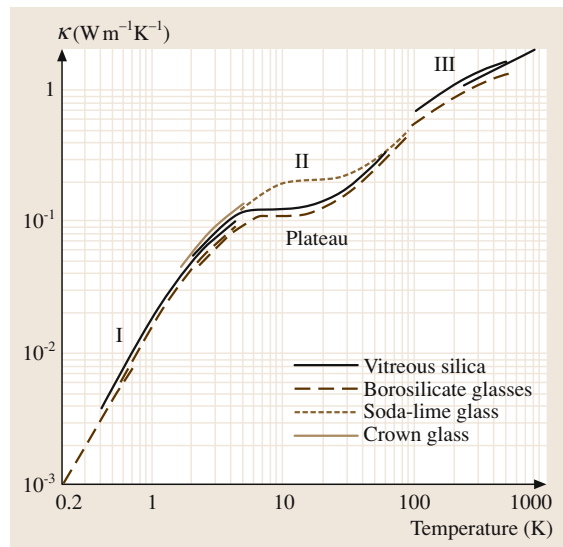


Fig. 19.8 The temperature dependence of κ for four types of glasses. Data in Figure 2 of Zeller and Pohl [19.21] replotted

ference has been qualitatively attributed to the phonon mean free path Λ_{ph} in a glass remaining roughly constant as the temperature is lowered and being limited by the range of disorder in the structure, which is a few atomic spacings [19.22, 23]. While intuitively this explanation seems reasonable, a more rigorous derivation excludes (19.17) since concepts like phonon velocity and wavelength tend to become meaningless over very short distances [19.24]. Most of the theories that have been proposed for describing κ more rigorously in the non-crystalline state are quite mathematical, and beyond the scope of this review based on materials science. (These models include the Kubo transport equation [19.25], hopping of localized vibrations [19.26], simulations based on molecular dynamics [19.27] etc.) Figure 19.8 shows κ versus T for a few selected glasses [19.21, 28]. Other glasses tend to follow this type of behavior in which there are three regions of distinct behavior. At very low temperatures (I in Fig. 19.8, below 1–2 K, $\kappa \approx T^n$ where $n = 1.8\text{--}2.0$). In the intermediate temperature regime, from about 2 K to about 20–30 K, κ exhibits a plateau. At high temperatures, above 20–30 K, beyond the plateau, κ increases with T , with a decreasing slope.

Polymers

Crystalline polymers tend to follow the general predictions in Fig. 19.7 that at high temperatures κ decreases with the temperature. As the degree of crystallinity decreases, κ also decreases, and eventually for amorphous polymers κ is substantially smaller than the crystalline phase, and κ increases with increasing temperature. If ρ_c and ρ_a are the densities of a given polymer in the crystalline (c) and amorphous (a) phase, then various empirical rules have been proposed to relate the conductivities of c- and a-phases through the ratio ρ_c/ρ_a of their densities. For example, in one empirical rule [19.29] $\kappa_c/\kappa_a = 1 + 1.58[(\rho_c/\rho_a) - 1]$. This means that, if we know κ_c , and ρ_c/ρ_a , we can calculate κ_a . If we know the density ρ of the partially crystalline polymer or the volume fraction v_a of the amorphous phase, then we can use an appropriate mixture rule to find the overall κ . (A solved problem is given as an example in chapter 17 of [19.10].) Figure 19.9 shows the dependence of κ on T for three different degrees of crystallinity in the structure; clearly κ versus T is strongly influenced by the crystallinity of the structure.

Figure 19.10 shows the temperature dependence of κ when it has been normalized to its value $\kappa(T_g)$ at the glass-transition temperature T_g against the normalized temperature, T/T_g for a number of amorphous

polymers. While there are marked differences between various polymers, there is nonetheless an overall trend in which κ increases up to T_g and thereafter decreases with T . The solid curve in Fig. 19.10 indicates an overall trend. Notice how similar the κ values are.

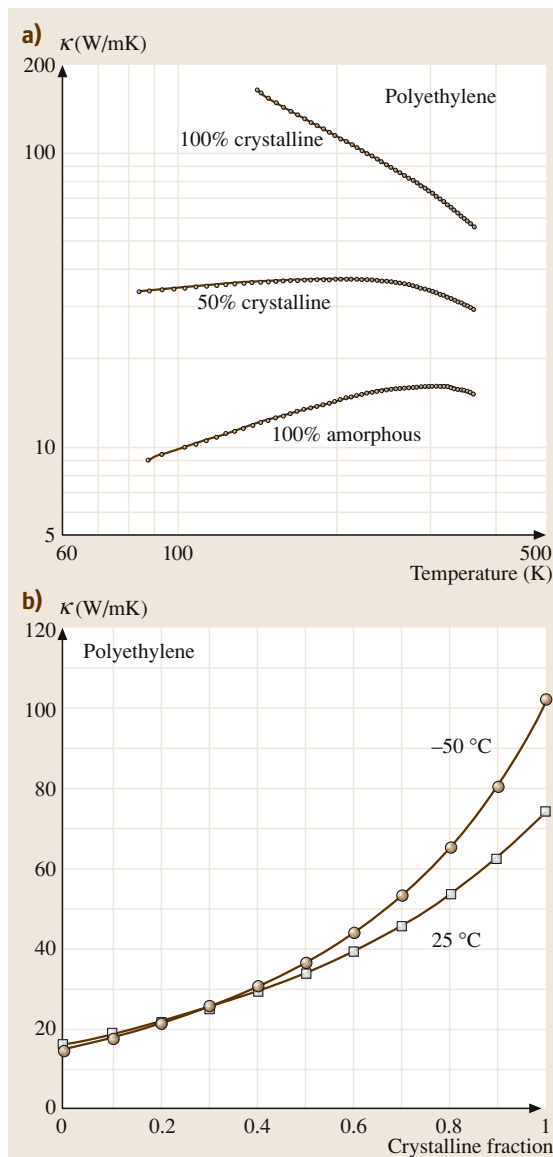


Fig. 19.9 (a) κ – T behavior for polyethylene for a fully crystalline, fully amorphous and 50% crystalline structures. (b) Dependence of κ on the crystallinity of polyethylene. Data extracted and replotted from [19.29]

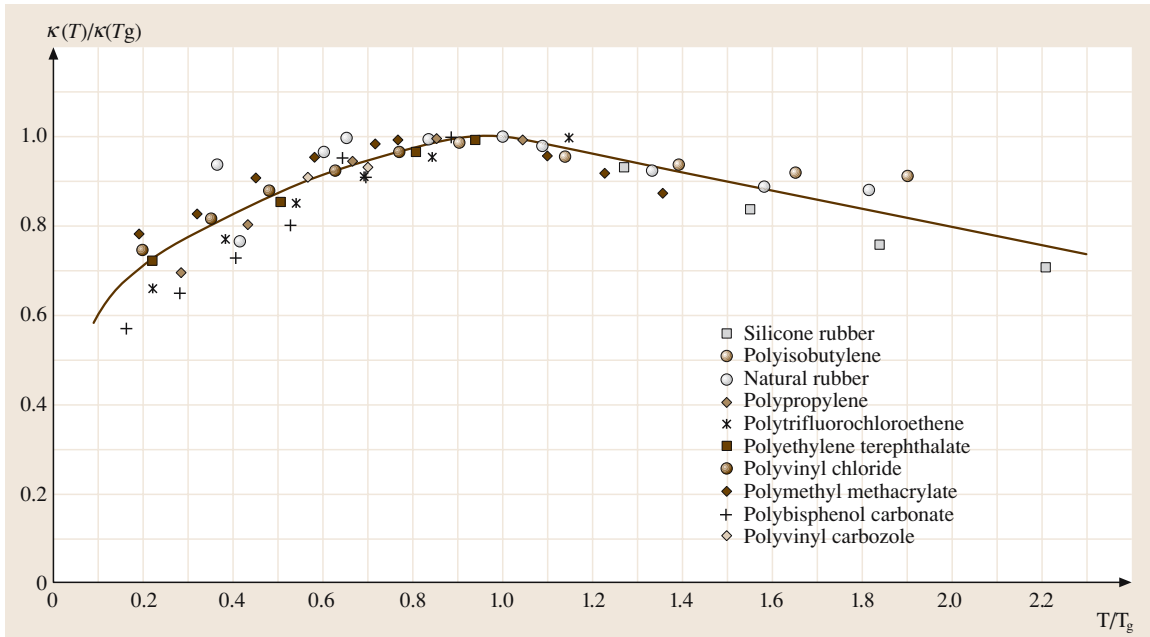


Fig. 19.10 Normalized thermal conductivity, $\kappa(T)/\kappa(T_g)$ versus normalized temperature, T/T_g , for various polymers. (Figure 17.2 in [19.10])

The overall or effective thermal conductivity of polymer blends can be usually estimated from the individual conductivities, the volume or weight fractions of the constituents, and the crystallinity of the structure; the problem can be quite complicated. A good example in which κ has been examined for a polymer blend as a function of blend composition has been given by Agari et al. [19.30], who also test various mixture rules.

The thermal conductivity of polymeric composite materials (as opposed to blends) can be estimated from various mixture rules [19.31, 32]. For example, if a matrix with κ_1 contains spherical particles with κ_2 , then the composite has a thermal conductivity,

$$\kappa = \kappa_1 \frac{2\kappa_1 + \kappa_2 - 2v_2(\kappa_1 - \kappa_2)}{2\kappa_1 + \kappa_2 + v_2(\kappa_1 - \kappa_2)}, \quad (19.19)$$

where v_2 is the volume fraction of the particles (with κ_2).

19.2.4 Thermal Conductivity of Metals

Although the heat capacity of metals is due the lattice vibrations, the thermal conductivity is due to the free conduction electrons. Both charge transport and heat transport are carried by the conduction electrons, and hence the electrical and thermal conductivity, σ and κ respectively, are related. Normally only electrons around

the Fermi level contribute to conduction and energy transport. In general κ and σ can be written as

$$\kappa = \frac{1}{3} C_e v_F \ell_e \quad \text{and} \quad \sigma = \frac{1}{3} e^2 v_F \ell_e g(E_F), \quad (19.20)$$

where C_e is the electronic heat capacity per unit volume, v_F is the Fermi velocity, $g(E_F)$ is the density of states at E_F (Fermi level), and ℓ_e is the electron mean free path. The scattering of electrons limits ℓ_e , which then limits κ and σ . If we then apply the free-electron theory to substitute for C_e and $g(E_F)$ then the ratio of κ to σT turns out to be a constant, which depends only on the constants k and e ; ℓ_e cancels out in this ratio. A proper derivation involves solving the Boltzmann transport equation in the presence of a temperature gradient [19.33], and leads to the well-known *Wiedemann–Franz–Lorenz* (WFL) law

$$\frac{\kappa}{\sigma T} = C_{\text{WFL}} = \frac{\pi^2 k^2}{3e^2} = 2.44 \times 10^{-8} \text{ W } \Omega / \text{K}^2, \quad (19.21)$$

where C_{WFL} is called the *Lorentz number*.

As shown in Fig. 19.11, the WFL law is well obeyed at room temperature for most pure metals and their alloys. The WFL law is normally well obeyed over a wide

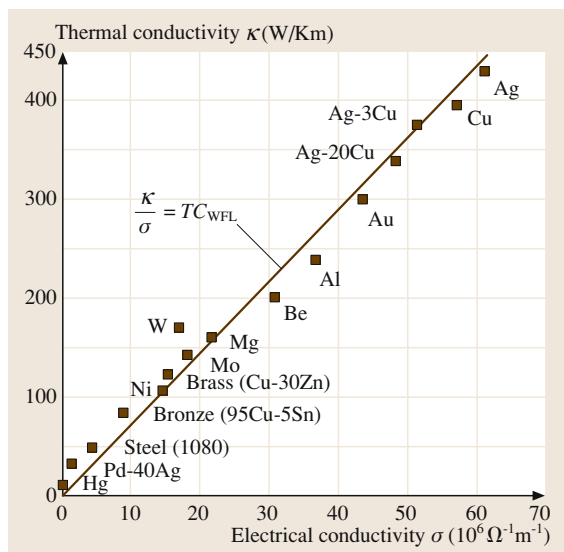


Fig. 19.11 Thermal conductivity versus electrical conductivity for various metals (both pure metals and alloys) at 20 °C. (From [19.5])

temperature range, and also by many liquid pure metals. Consequently the thermal conductivity κ of a metal or an alloy can be readily calculated given the conductivity or resistivity ρ . For example, we can easily calculate the resistivity of a solid solution alloy by using Matthiessen's and Nordheim's rules (as explained in this handbook by Koughia et al. in Chapt. 2) and then simply calculate κ from (19.21). Figure 19.12 shows the typical dependence of κ on T for two pure metals (Cu and Al) and two alloys (brass and Al-14%Mg). Above the Debye temperature, for pure metals, $\sigma \propto 1/T$, so that the thermal conductivity remains almost constant, as in Fig. 19.12. For alloys, σ values are smaller than the corresponding pure metals and σ exhibits a weaker temperature depen-

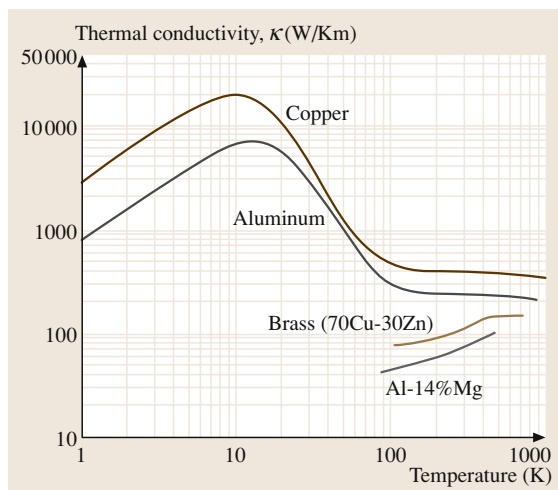


Fig. 19.12 Thermal conductivity versus temperature for two pure metals (Cu and Al) and two alloys (brass and Al-14% Mg). (From [19.5])

dence. Consequently κ is smaller and increases with the temperature, as shown in Fig. 19.12.

While for pure metals and many alloys κ is due to the conduction electrons, the phonon contribution can nonetheless become comparable to the electron contribution for those alloys that have substantial resistivity or at low temperatures. Glassy metals (or metallic glasses) usually have two or more elements alloyed; one of the elements is usually a nonmetal or a metalloid. Their resistivities tend to be higher than usual crystalline or polycrystalline alloys. Both conduction electrons and also phonons contribute to the thermal conductivity of these glassy metals. For example, for $\text{Fe}_{80}\text{B}_{20}$, $\rho = 1220 \text{ n}\Omega \text{ m}$ (compare with $17 \text{ n}\Omega \text{ m}$ for Cu), and the phonon contribution is about 15% at 273 K and 61% at 100 K [19.34].

19.3 Thermal Expansion

19.3.1 Grüneisen's Law and Anharmonicity

Nearly all materials expand as the temperature increases. The thermal coefficient of linear expansion, or simply, the thermal expansion coefficient α is defined as the fractional change in length per unit temperature increase, or

$$\alpha = \frac{1}{L_0} \cdot \left(\frac{dL}{dT} \right)_{T_0}, \quad (19.22)$$

where L_0 is the original length of the substance at temperature T_0 , dL/dT is the rate of change of length with temperature at the reference temperature T_0 .

The principle of thermal expansion is illustrated in Fig. 19.13, which is the potential-energy curve $U(r)$ for two atoms separated by a distance r in a crystal. At temperature T_1 , the atoms would be vibrating about their equilibrium positions between B and C ; compressing (B) and stretching (C) the bond between them. The line BC corresponds to the total energy E of the pair

Table 19.3 The Grüneisen parameter for some selected materials with different types of interatomic bonding (FCC = face-centered cubic; BCC = body-centered cubic)

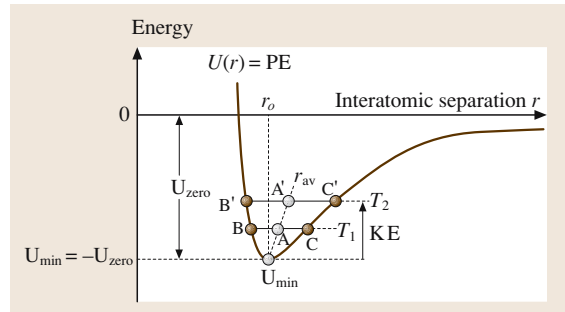
Material	ρ (g/cm ³)	α ($\times 10^{-6}$ K ⁻¹)	K (GPa)	c_s (J/kg K)	γ
Iron (metallic, BCC)	7.9	12.1	170	444	0.20
Copper (metallic, FCC)	8.96	17	140	380	0.23
Germanium (covalent)	5.32	6	77	322	0.09
Glass (covalent-ionic)	2.45	8	70	800	0.10
NaCl (ionic)	2.16	39.5	28	880	0.19
Tellurium (mixed)	6.24	18.2	40	202	0.19
Polystyrene (van der Waals)	1.05	100	3	1200	0.08

of atoms. The average separation at T_1 is at A, halfway between B and C. The PE curve $U(r)$ is asymmetric, and it is this asymmetry that leads to the phenomenon of thermal expansion. When the temperature increases from T_1 to T_2 the atoms vibrate between B' and C' and the average separation between the atoms also increases, from A to A', which is identified as thermal expansion. If the PE curve were symmetric, then there would be no thermal expansion. The extent of expansion (A to A') depends on the amount of increase from BC to B'C' per degree of increase in the temperature. α must therefore also depend on the heat capacity. When the temperature increases by a small amount δT , the energy per atom increases by $(C_v \delta T)/N$, where C_v is the heat capacity per unit volume and N is the number of atoms per unit volume. If $C_v \delta T$ is large then the line B'C' in Fig. 19.13 will be higher up on the energy curve and the average separation A' will therefore be larger. Thus, $\alpha \propto C_v$. Further, the average separation, point A, depends on how much the bonds are stretched and compressed. For large amounts of displacement from equilibrium, the average A will be greater as more asymmetry of the PE curve is used. Thus, smaller is the elastic modulus K , the greater is α ; clearly $\alpha \propto C_v/K$.

If we were to expand $U(r)$ about its minimum value U_{\min} at $r = r_0$, we would obtain the Taylor expansion to the cubic term as,

$$U(r) = U_{\min} + (1/2)\beta(r - r_0)^2 - (1/3)g(r - r_0)^3, \quad (19.23)$$

where β and g are coefficients related to second and third derivatives of U at r_0 . The term $(r - r_0)$ is missing because we are expanding a series about U_{\min} where $dU/dr = 0$. The U_{\min} and the $\beta_2(r - r_0)^2$ term give a parabola about U_{\min} which is a symmetric curve around r_0 and therefore does not lead to thermal expansion. It is the cubic term that gives the expansion because it leads to asymmetry. Thus, the amount of expansion α

**Fig. 19.13** Interatomic potential energy (PE) U curve, which has a minimum when the atoms in the solid attain the interatomic separation at $r = r_0$

also depends on the amount of asymmetry with respect to symmetry that is β/g . Thus, α is proportional to both C_v/K and g/β . The third term in (19.23) is called *anharmonicity* in the interatomic PE curve. The ratio of g to β depends on the nature of the bond. A simplified analytical treatment [19.35] gives α as,

$$\alpha \approx 3\gamma \frac{C_v}{K}; \quad \gamma = \frac{r_0 g}{3\beta}, \quad (19.24)$$

where γ is a constant called the *Grüneisen parameter*, which represents the relative asymmetry of the energy curve, g/β . γ is of the order of unity for many materials; experimentally, $\gamma = 0.1$ – 1 . We can also write the Grüneisen law in terms of the molar heat capacity C_m [heat capacity per mole or the specific heat capacity c_s (heat capacity per unit mass)]. If ρ is the density, and M_{at} is the atomic mass of the constituent atoms of the crystal, then

$$\lambda = 3\gamma \frac{\rho C_m}{M_{\text{at}} K} = 3\gamma \frac{\rho c_s}{K}. \quad (19.25)$$

Given the experimental values for α , K , ρ and c_s , the Grüneisen parameters have been calculated from (19.25)

and are listed in Table 19.3. An interesting feature is that the experimental γ values, within a factor of 2–3, are about the same, at least to an order of magnitude. Equation (19.24) also indicates that the dependence of α on T should resemble the dependence of C_v on T , which is borne out by experiments.

19.3.2 Thermal Expansion Coefficient α

The thermal expansion coefficient normally depends on the temperature, $\alpha = \alpha(T)$, and typically increases with increasing temperature, except at the lowest temperatures. Figure 19.14 shows α versus T for selected materials and also compares α for a wide range of materials at 300 K. In very general terms, except at very low (typically below 100 K) and very high temperatures (near the melting temperature), for most metals α does not depend strongly on the temperature; many engineers take α for a metal to be approximately temperature independent. Further, higher-melting-point temperature metals tend to have lower α .

To calculate the final length L from the original length L_0 that corresponds to the temperature increasing from T_0 to T , we have to integrate $\alpha(T)$ with respect to temperature from T_0 to T . If $\bar{\alpha}$ is the mean value for the expansion coefficient from T_0 to T , then

$$L = L_0[1 + \bar{\alpha}(T - T_0)] \text{ and} \quad (19.26)$$

$$\bar{\alpha} = \frac{1}{(T - T_0)} \int_{T_0}^T \alpha(T) dT \quad (19.27)$$

For example, the expansion coefficient of silicon over the temperature range 120–1500 K is given by [19.36]

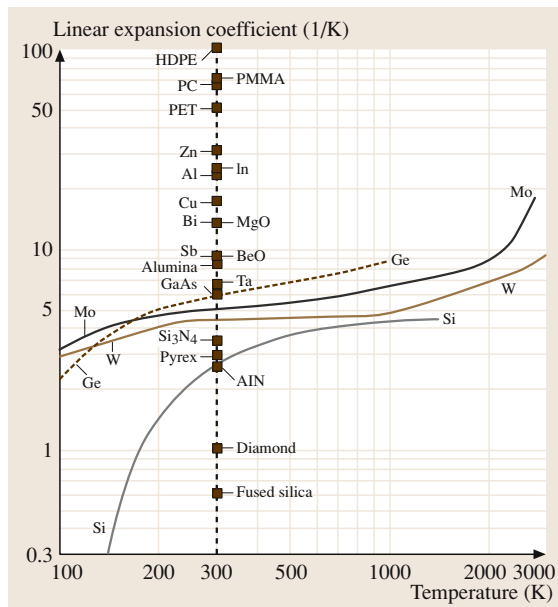


Fig. 19.14 α versus T for selected materials and comparison of α for a range of materials at 300 K. (After [19.5])

as

$$\alpha = 3.725 \times 10^{-6} \left(1 - e^{-5.88 \times 10^{-3}(T-124)} \right) + 5.548 \times 10^{-10} T, \quad (19.28)$$

where T is in Kelvin. At 20 °C, $\alpha = 2.51 \times 10^{-6} \text{ K}^{-1}$, whereas from 20 °C to 320 °C, $\bar{\alpha} = 3.35 \times 10^{-6} \text{ K}^{-1}$. Most $\alpha(T)$ expressions have exponential-type expressions as in (19.28) but they can always be expanded in terms of T or $(T - T_0)$ up to the third term, beyond which errors become insignificant.

19.4 Enthalpic Thermal Properties

19.4.1 Enthalpy, Heat Capacity and Physical Transformations

Consider the behavior of the enthalpy $H(T)$ versus temperature T of a typical material as it is cooled from a liquid state starting at O as shown in Fig. 19.15 (the sketch is idealized to highlight the fundamentals). For crystalline solids, at the melting or fusion temperature T_m , the liquid L solidifies to form a crystalline solid C . Some degree of undercooling is necessary to nucleate the crystals but we have ignored this fact

in Fig. 19.15 so that T_m is the thermodynamic melting temperature where the Gibbs free energies of the liquid (L) and the crystal (C) are the same. The heat capacity $C_{\text{liquid}} = dH_{\text{liquid}}/dT$ for the liquid is greater than that for the crystal, $C_{\text{crystal}} = dH_{\text{crystal}}/dT$, as apparent from the H - T slopes in Fig. 19.15. The fusion enthalpy ΔH_f corresponds to the enthalpy change between the liquid and crystalline phases, L and C , in Fig. 19.15.

When a glass-forming liquid is cooled, it does not crystallize at T_m ; its crystallization kinetics are too slow or negligible. Its enthalpy H continues to decrease along

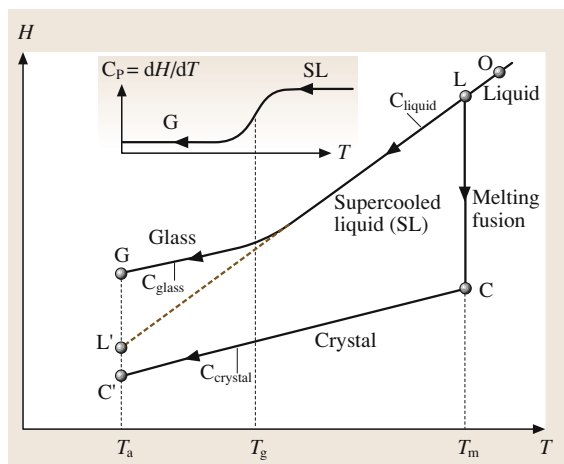


Fig. 19.15 Enthalpy H versus temperature T

the liquid H versus T line with a slope C_{liquid} and it becomes a supercooled liquid. However, as the cooling proceeds, eventually, over a certain small temperature range around a temperature marked as T_g in Fig. 19.15, the atomic motions become too sluggish, and the viscosity becomes too high to follow the liquid-like behavior. The material cools along a solid-like H - T curve that is almost parallel to the crystalline or solid H - T line. This region where the slope or the heat capacity is similar to that of the solid or the crystal (C_{crystal}) is called the glassy state. The intersection of the liquid-like (supercooled) and solid-like (glass) H - T lines defines a *glass-transition temperature* T_g as observed under cooling at a particular cooling rate q . T_g depends on the cooling rate, and obviously on the structure. C_p changes in a step-like fashion over a small interval around T_g during cooling as shown in Fig. 19.15. The shift in T_g with q can provide valuable information on the kinetics of the glass-transformation process as discussed below. The glass enthalpy is greater than the crystalline enthalpy due to the disorder in the glass structure.

Suppose we stop the cooling at an ambient temperature T_a , corresponding to the glass state G . The glass enthalpy $H_{\text{glass}}(T_a)$ will relax via structural relaxations toward the enthalpy of the metastable liquid-like equilibrium state $H_{\text{liquid}}(T_a)$ at L' . This process is called aging or annealing. When we start heating a glass, its initial enthalpy can be anywhere between G and L' , depending on the thermal history (including aging) and the cooling rate.

When the sample at $H(T_a)$ at G' is heated at a constant rate r , as shown in Fig. 19.16, the enthalpy of

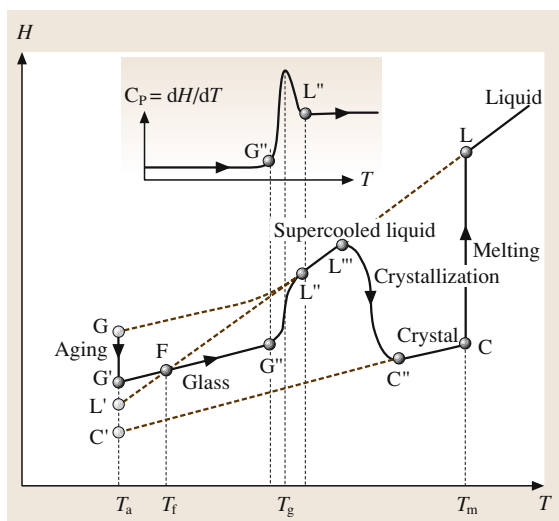


Fig. 19.16 Isothermal aging at T_a changes the glass enthalpy. When the glass is heated at a constant rate r , it undergoes glass transition at around T_g , and then ideally crystallization and then melting

the system, as a result of long structural relaxation times at these low temperatures, follows the glass H - T curve until the structural relaxation rate is sufficiently rapid to allow the system to recover toward equilibrium $H_{\text{liquid}}(T)$; the transition is from G'' to L'' . The sigmoidal-like change in $H(T)$ leads to the glass-transformation endotherm observed for many glasses when they are heated through T_g . This endothermic heat, called the enthalpy of relaxation, depends on the duration of aging, and the structural relaxation processes, which in turn depend on the structure (e.g. [19.37]). The observed glass-transition temperature T_g during heating depends not only on the heating rate r but also on the initial state $H(T_a)$ or G' . On the other hand, during cooling from the melt down to low temperatures, one observes only a change in the heat capacity. As we further heat the sample, the supercooled liquid state can crystallize, from L''' to C'' and form a true solid, usually polycrystalline. This transformation from L''' to C'' releases heat; it is exothermic. At T_m , the solid melts to form the liquid, C changes to L , which means that heat is absorbed.

It is apparent from the simple H - T behavior in Fig. 19.16 that there are at least several key parameters of interest in characterizing the thermal properties of materials: the heat capacity of the liquid, crystal and glass phases, C_{liquid} , C_{crystal} , C_{glass} , their temperature dependences, melting or fusion temperature T_m , fusion enthalpy ΔH_f , crystallization onset temperature (L'''),

exothermic enthalpy of crystallization ($L''' - C'''$), T_g (which depends on various factors), and the endothermic glass-transformation enthalpy ($L'' - G''$). These properties are most conveniently examined using a *differential scanning calorimeter* (DSC).

19.4.2 Conventional Differential Scanning Calorimetry (DSC)

Principles

Differential scanning calorimetry (DSC) is a method which is extensively used to measure heats and temperatures of various transitions and has been recognized as a very useful tool for the interpretation of thermal events as discussed and reviewed by a number of authors (see for example [19.38–44]). In essence, the DSC measures the net rate of heat flow $Q' = dQ/dT$, into a sample with respect to heat flow into a reference inert sample, as the sample temperature is ramped. Figure 19.17 shows a simplified schematic diagram of the heat-flux DSC cell. The constantan disk allows heat transfer to and from the reference and sample pans by thermal conduction. Due to symmetry, the thermal resistance of the heat paths from the heater to the reference and sample pans is approximately the same. The constantan disk has two slightly raised platforms on which the reference and sample pans are placed. Chromel disks fixed under these platforms form area thermocouple junctions with the constantan platforms. The constantan-chromel thermocouples are in series and read the temperature difference ΔT between the reference and sample pans. There are also alumel–chromel thermocouples attached to the chromel disks to measure the individual reference and sample temperatures.

The measured temperature difference ΔT corresponds to a differential heat flow $Q' = dQ/dt$ between the sample and the reference pans by virtue of Fourier's

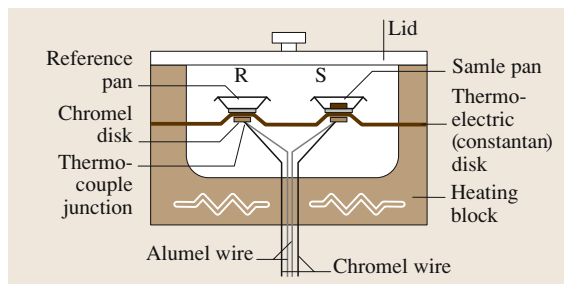


Fig. 19.17 Schematic illustration of heat-flux DSC cell. R and S refer to reference and sample pans

law of heat conduction,

$$Q' = \Delta T/R, \quad (19.29)$$

where R is the effective thermal resistance of the heat flow path to the sample (or reference) through the constantan disk. In conventional thermal analysis, the sample temperature is either ramped linearly at a constant heating or cooling rate or kept constant (as in isothermal experiments).

A typical example of a conventional DSC is shown in Fig. 19.18 where the thermogram evinces three distinct regions that correspond to the glass transition (T_g), crystallization and melting. T_g in DSC analysis is normally defined in terms of an onset temperature as in Fig. 19.18. The crystallization exotherm is normally used to obtain a crystallization onset temperature T_c and a peak crystallization rate temperature T_p that corresponds to the location of the peak crystallization. ΔH_c and fusion ΔH_f enthalpies are obtained by integrating the corresponding peaks as shown in Fig. 19.18 with an appropriate baseline. The base line represents the overall heat capacity of the sample, which changes during the transformation since the sample is gradually transformed into the new phase. Usually a simple sigmoidal, or even a straight line between the onset and the end points, is sufficient for an accurate determination of the enthalpy.

Conventional DSC has been used to study and characterize a wide range of physical and chemical transformations. When the DSC is used in a heating or cooling scan, the temperature of the sample is ramped along

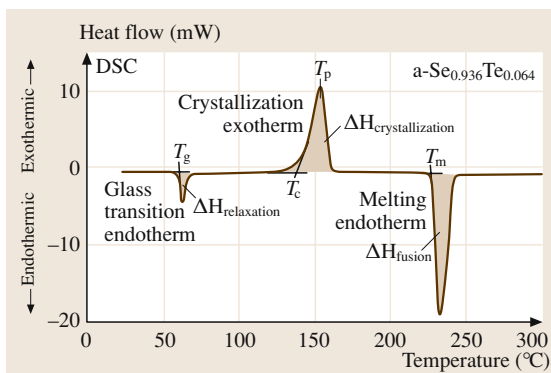


Fig. 19.18 Typical DSC signal for a chalcogenide $\text{Se}_{0.936}\text{Te}_{0.064}$ glass sample (well relaxed or aged). From left to right: the glass transition, crystallization and the melting temperatures and corresponding enthalpies of transitions for this material. $\Delta H_c = \Delta H_{\text{crystallization}}$; $\Delta H_f = \Delta H_{\text{fusion}}$

$T = T_0 + rt$, where T_0 is the initial temperature. Consequently, the transformations in DSC are carried out under non-isothermal conditions, and well-known isothermal rate equations cannot be directly applied without some modification.

Glass Transformation

There are extensive discussions in the literature on the meaning of the glass-transition region and the corresponding T_g (e.g. [19.45–50]). The most popular interpretation of T_g is based on the fact that this transformation is a kinetic phenomenon. The glass-transformation kinetics have been most widely studied by examining the shift in T_g with the heating or cooling rate in a so-called T_g -shift technique. The relaxation process can be modeled by assuming that the glass structure has a characteristic structural relaxation time that controls the rate at which the enthalpy can change. It is well recognized that the glass-transformation kinetics of glasses are nonlinear. In the simplest description, the relaxation can be conveniently described by using a single phenomenological relaxation time τ (called the Narayanaswamy or Tool–Narayanaswamy–Moynihan relaxation time) that depends not only on the temperature but also on the glass structure through the fictive temperature T_f as

$$\tau(T, T_f) = \tau_0 \exp[x\Delta h^*/RT + (1-x)\Delta h^*/RT_f], \quad (19.30)$$

where Δh^* is the activation enthalpy, T_f is the fictive temperature and x is the partition parameter which determines the relative contributions of temperature and structure to the relaxation process. T_f is defined in Fig. 19.16 as the intersection of the glass line passing through the starting enthalpic state G' and the extended liquid H – T lines. It depends on the starting enthalpy G' so that T_f is used as a convenient temperature parameter to identify the initial state at G' . Due to the presence of the structural parameter x , the activation energies obtained by examining the heating and cooling rate dependences of the glass-transition temperature are not the same. If the shift in T_g is examined as a function of the cooling rate q starting from a liquid-like state (above T_g) then a plot of $\ln q$ versus $1/T_g$ (called a *Ritland plot* [19.51]) should yield the activation enthalpy Δh^* in (19.30) [19.52–54]. In many material systems, the relaxation time τ in (19.30) is proportional to the viscosity, $\tau \propto \eta$ [19.55–57] so that Δh^* from cooling scans agrees with the activation energy for the viscosity [19.58–62] over the same temperature range. The viscosity η usually

follows either an Arrhenius temperature dependence, as in oxide glasses, or a Vogel–Tammann–Fulcher behavior, $\eta \propto \exp[A/(T - T_0)]$, where A and T_0 are constants, as in many polymers and some glasses, e.g. chalcogenides.

The relaxation kinetics of various structural properties such as the enthalpy, specific volume, elastic modulus, dielectric constant etc. have been extensively studied near and around T_g , and there are various reviews on the topic (e.g. [19.63]). One particular relaxation kinetics that has found widespread use is the stretched exponential in which the rate of relaxation of the measured property is given by

$$\text{Rate of relaxation} \propto \exp\left[-\left(\frac{t}{\tau}\right)^\beta\right], \quad (19.31)$$

where β (< 1) is a constant that characterizes the departure from the pure exponential relaxation rate. Equation (19.31) is often referred to as the *Kohlrausch–Williams–Watts* (KWW) [19.64] stretch exponential relaxation function. β depends not only on the material but also on the property that is being studied. In some relaxation processes, the whole relaxation process over a very long time is sometimes described by two stretched exponentials to handle the different fast and slow kinetic processes that take place in the structure [19.65].

The kinetic interpretation of T_g implies that, as the cooling rate is slowed, the transition at T_g from the supercooled liquid to the glass state is observed at lower temperatures. There is however a theoretical thermodynamic boundary to the lowest value of T_g . As the supercooled liquid is cooled, its entropy decreases faster than that of the corresponding crystal because $C_{\text{liquid}} > C_{\text{crystal}}$. Eventually at a certain temperature T_0 , the relative entropy lost $\Delta S_{\text{liquid-crystal}}$ by the supercooled liquid with respect to the crystal will be the same as the entropy decrease (latent entropy of fusion) $\Delta S_f = \Delta H_f/T_m$ during fusion. This is called *Kauzmann's paradox* [19.66], and the temperature at which $\Delta S_{\text{liquid-crystal}} = \Delta S_f$ is the lowest theoretical boundary for the glass transformation; $T_g > T_0$.

The changes in T_g with practically usable heating or cooling rates are usually of the order of 10 °C or so. There have been various empirical rules that relate T_g to the melting temperature T_m and the glass structure and composition. Since T_g depends on the heating or cooling rate, such rules should be used as an approximation; nonetheless, they are extremely useful in engineering as a guide to the selection and use of materials.

Table 19.4 Some selected examples of T_g dependences on various factors

Rule	Notation	Comment
$T_g \approx (2/3)T_m$	T_g and T_m in K. T_m = melting temperature of corresponding crystalline phase.	Kauzmann's empirical rule [19.66]. Most glass structures including many amorphous polymers [19.67]. Some highly symmetrical polymers with short repeat units follow $T_g \approx (1/2)T_m$ [19.10]
$\ln(q) \approx -\Delta h^*/RT_g + C$	q = cooling rate; Δh^* = activation energy in (19.30); C = constant.	Dependence of T_g on the cooling rate. Δh^* may depend on the range of temperature accessed. Bartenev–Lukianov equation [19.51, 68]
$T_g \approx T_g(\infty) + C/M_n$	M_n = average molecular weight of polymer; C = constant; $T_g(\infty)$ is T_g for very large M_n	Dependence of T_g on the average molecular weight of a polymer [19.69, 70]. Tanaka's rule
$\ln(T_g) \approx 1.6Z + C$	Z = mean coordination number, C = constant (≈ 2.3)	Network glasses. Dependence of T_g on the mean coordination number. Neglects the heating rate dependence. [19.71]
$T_g(x) = T_g(0) - 626x$	T_g in K; x is atomic fraction in $a: (\text{Na}_2\text{O} + \text{MgO})_x(\text{Al}_2\text{O}_3 + \text{SiO}_2)_{1-x}$ $b: (\text{PbO})_x(\text{SiO}_2)_{1-x}$ $c: (\text{Na}_2\text{O})_x(\text{SiO}_2)_{1-x}$ $T_g(0) = 1080 \text{ K for } a;$ $967 \text{ K for } b; 895 \text{ K for } c.$	$\pm 5\%$. Silicate glasses [19.72]; x is network modifier. $0.01 < x < 0.6$

Non-Isothermal Phase Transformations

The crystallization process observed during a DSC heating scan is a non-isothermal transformation in which nucleation and growth occur either at the same time as in homogenous nucleation or nucleation occurs before growth as in heterogenous nucleation. In the case of isothermal transformations by nucleation and growth, the key equation is the so-called *Johnson–Mehl–Avrami equation*,

$$x = 1 - \exp(-Kt^n), \quad (19.32)$$

where $K \propto \exp(-E_A/k_B T)$ is the thermally activated rate constant, and n is a constant called the *Avrami index* whose value depends on whether the nucleation is heterogeneous or homogenous, and the dimensionality m of growth ($m = 1, 2$ or 3 for one-, two- or three-dimensional growth). For example, for growth from preexisting nuclei (heterogeneous nucleation) $n = m$, and for continuing nucleation during growth, $n = m + 1$. A detailed summary of possible n and m values has been given by Donald [19.73]

DSC studies however are conventionally non-isothermal. There have been numerous papers and discussions on how to extract the kinetic parameters of the transformation from a DSC non-isothermal experiment [19.74–78]. It is possible to carry out a reasonable examination of the crystallization kinetics by combining a single scan experiment with a set of multiple scans; there are many examples in the literature (e.g. [19.78, 79]). Suppose that we take a single DSC

scan, as in Fig. 19.18, and calculate the fraction of crystallized material x at a temperature T . The plot of $\ln[-\ln(1-x)]$ versus $1/T$ (*Coats–Redfern–Sestak plot* [19.80, 81]) then provides an activation energy E'_A from a single scan. For heterogenous nucleation, E'_A is mE_G , whereas for homogenous nucleation it is $E_N + mE_G$, where E_G is the activation energy for growth and E_N is the activation energy for nucleation. Clearly, we need to know something about the nucleation process and dimensionality of growth to make a sensible use of E'_A . Thus,

$$\ln[-\ln(1-x)] = -\frac{E'_A}{RT} + C', \quad (19.33)$$

where C' is a constant.

Suppose we then examine how the peak rate temperature T_p shifts with the heating rate r . Then a plot of $\ln(r/T_p^2)$ versus $1/T_p$ is called a *Kissinger plot* [19.82, 83], and gives an activation energy E''_A . Thus,

$$\ln\left(\frac{r}{T_p^2}\right) = -\frac{E''_A}{RT_p} + C'', \quad (19.34)$$

where C'' is a constant. In heterogenous nucleation E''_A simply represents the activation energy of growth E_G , whereas if the nucleation continues during growth, $E''_A = (E_N + mE_G)/(m+1)$. The ratio E'_A/E''_A represents the non-isothermal Avrami index n . Table 19.5 provides an overview of various thermal analysis techniques that have been used for characterizing non-isothermal phase transformations.

Table 19.5 Typical examples of studies of transformation kinetics. Usual interpretation of E'_A and E''_A are $E'_A = mE_G$ or $(E_N + mE_G)$ and $E''_A = E_G$ or $(E_N + mE_G)/(m + 1)$; \dot{x} is the rate of crystallization

Study	Method and plot	Slope provides	Method
Single scan	$\ln[-\ln(1-x)]$ versus $1/T$	E'_A	<i>Coats-Redfern-Sestak</i> [19.80,81]
Single scan	$\frac{\dot{x}}{(1-x)[-\ln(1-x)]^{(n-1)/n}}$	E''_A	If n chosen correctly, this agrees with the Kissinger method [19.77,79]. Independent of the initial temperature
Multiple scan	$\ln(r/T_p^2)$ versus $1/T_p$	E''_A	<i>Kissinger</i> [19.82,83]. Initial temperature effect in [19.77]
Multiple scan	$\ln(r^n/T_p^2)$ versus $1/T_p$	mE''_A	Modified Kissinger; <i>Matusita, Sakka</i> [19.89,90]
Multiple scan	$\ln r$ versus $1/T_c$	E''_A	<i>Ozawa method</i> [19.74–76]
Multiple scan	$\ln[r/(T_p - T_c)]$ versus $1/T_p$; T_c = initial temperature	E''_A	<i>Augis, Bennett</i> [19.91]
Multiple scan	$[d\Delta H/dt]_{\max}$ versus $1/T_p$; $\dot{x} = [d\Delta H/dt]_{\max}$	E''_A	<i>Borchardt-Pilonyan</i> [19.92,93]
Multiple scan	$\ln(r/T_1^2)$ versus $1/T_1$; when $x = x_1$, $T = T_1$; T_1 depends on r	E''_A	<i>Ozawa-Chen</i> [19.94,95]
Multiple scan	$\ln[-\ln(1-x)]$ versus $\ln r$; x is the crystallized amount at $T = T_1$; T_1 is constant	$-n$	<i>Ozawa method</i> [19.90,94]

DSC has been widely used to study the kinetics of crystallization and various phase transformations occurring in a wide range of material systems; there are numerous recent examples in the literature [19.84–88]. Equations (19.33) and (19.34) represent a simplified analysis. As emphasized recently [19.73], a modified Kissinger analysis [19.89,90] involves plotting $\ln(r^n/T_p^2)$ versus $1/T_p$, the slope of which represents an activation energy, mE''_A ; however, the latter requires

some knowledge of n or m to render the analysis useful. n can be obtained examining the dependence of $\ln[-\ln(1-x)]$ on $\ln r$ at one particular temperature, which is called the Ozawa method as listed in Table 19.5 [19.90]. It is possible to combine the modified Kissinger analysis with an isothermal study of crystallization kinetics to infer n and m given the type of nucleation process (heterogeneous or homogenous) that takes place.

19.5 Temperature-Modulated DSC (TMDSC)

19.5.1 TMDSC Principles

In the early 1990s, a greatly enhanced version of the DSC method called temperature-modulated differential scanning calorimetry (MDSC™) was introduced by the efforts of *Reading* and coworkers [19.96–98]. The MDSC method incorporates not only the ability of conventional DSC but it also provides significant and distinct advantages over traditional DSC. The benefits of the MDSC technique have been documented in several recent papers, and include the following: separation of complex transitions, e.g. glass transition, into easily interpreted components; measurement of heat flow and heat capacity in a single experiment; ability to determine more accurately the initial crystallinity of the studied material; increased sensitivity

for the detection of weak transitions; increased resolution without the loss of sensitivity; measurement of thermal conductivity [19.99,100]. One of the most important benefits is the separation of complex transitions such as the glass transition into more easily interpreted components. Recent applications of MDSC to glasses has shown that it can be very useful for the interpretation of thermal properties, such as the heat capacity, in relation to the structure as, for example, in the case of chalcogenide glasses (e.g. [19.15,101,102]).

The MDSC™ that is currently commercialized by TA Instruments uses a conventional heat-flux DSC cell whose heating block temperature is sinusoidally modulated. In MDSC, the sample temperature is modulated sinusoidally about a constant ramp so that the tempera-

ture T at time t is,

$$T = T_0 + rt + A \sin(\omega t), \quad (19.35)$$

where T_0 is the initial (or starting) temperature, r is the heating rate (which may also be a cooling ramp q), A is the amplitude of the temperature modulation, $\omega = 2\pi/P$ is the angular frequency of modulation and P is the modulation period. It should be emphasized that (19.35) is a simplified statement of the fact that the cell has reached a steady-state operation and that the initial temperature transients have died out. The resulting instantaneous heating rate, dT/dt , therefore varies sinusoidally about the average heating rate r , and is given by

$$dT/dt = r + A\omega \sin(\omega t). \quad (19.36)$$

At any time, the apparatus measures the sample temperature and the amplitude of the instantaneous heat flow (by measuring ΔT) and then, by carrying out a suitable Fourier deconvolution of the measured quantities, it determines two quantities (which have been termed by TA Instruments):

1. Reversing heat flow (RHF),
2. Nonreversing heat flow (NHF).

Fourier transforms are made on one full cycle of temperature variation, which means that the average quantities refer to moving averages. The average heat flow, which corresponds to the average heating rate (r), is called the total heat flow (HF). Total heat flow is the only quantity that is available and hence it is the only quantity that is always measured in conventional DSC experiments. MDSC determines the heat capacity using the magnitudes of heat flow and heating rate obtained by averaging over one full temperature cycle. If triangular brackets are used for averages over one period P , then $\langle Q' \rangle$ is the average heat flow per temperature cycle. The heat capacity per cycle is then calculated from

$$mC_P = \langle \text{Heat flow} \rangle / \langle \text{Heating rate} \rangle, \quad (19.37)$$

where m is the mass of the sample. This C_P has been called the *reversing heat capacity*, though Schawe has defined as the complex heat capacity [19.12]. The reversing in this context refers to a heat flow that is reversing over the time scale of the modulation period. Furthermore, it is assumed that C_P is constant, that is, it does not change with time or temperature over the modulation period. The reversing heat flow is then obtained by

$$\text{RHF} = C_P \langle dT/dt \rangle. \quad (19.38)$$

The nonreversing heat flow (NHF) is the difference between the total heat flow and the reversing heat flow and represents heat flow due to a kinetically hindered process such as crystallization. There are a few subtle issues in the above condensed qualitative explanation. First is that the method requires several temperature cycles during a phase transition to obtain RHF and NHF components, which sets certain requirements on r , A and ω . The second is that the phase difference between the heat flow and the heating rate oscillations is assumed to be small as it would be the case through a glass-transition region or crystallization; but not through a melting process. There are a number of useful discussions and reviews on the MDSC technique and its applications in the literature [19.103–107].

19.5.2 TMDSC Applications

At present there is considerable scientific interest in applying TMDSC measurements to the study of glass-transformation kinetics in glasses and polymers (e.g. [19.15, 101, 108–112]). The interpretation of TMDSC measurements in the glass-transition region has been recently discussed and reviewed by *Hutchinson* and *Montserrat* [19.113–116]. The reversing heat flow (RHF) through the T_g region exhibits a step-like change and represents the change in the heat capacity. The hysteresis effects associated with thermal history seem to be less important in the RHF but present in the NHF. The measurement of T_g from the RHF in TMDSC experiments shows only a weak dependence on glass aging and thermal history [19.117–119], which is a distinct advantage of this technique. The interpretation of the NHF has been more difficult but it is believed that it provides a qualitative indication of the enthalpy loss during the annealing period below T_g [19.103] though more research is needed to clarify its interpretation. Figure 19.19 shows a typical MDSC result through the glass-transition region of $\text{Se}_{99.5}\text{As}_{0.5}$ glass.

It is important to emphasize that ideally the underlying heating rate in TMDSC experiments should be as small as possible. In this way we can separate the conventional DSC experiment, which also takes place during TMDSC measurements, from the dynamic, frequency-controlled TMDSC experiment. The oscillation amplitude A in the TMDSC must be properly chosen so that the C_P measurements do not depend on A ; typically $A = \pm 1.0^\circ\text{C}$ [19.61]. The oscillation period P should be chosen to ensure that there are at least four full modulations within the half width of the temperature transition, that is, a minimum of eight oscillations over

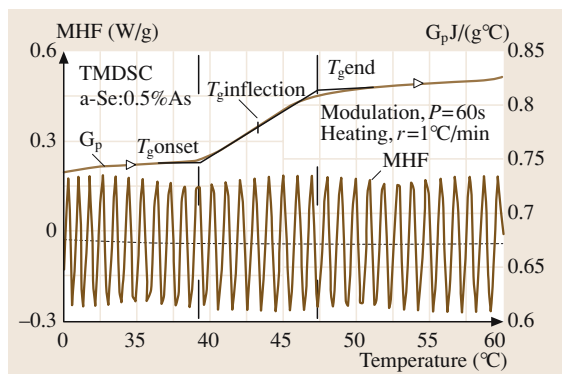


Fig. 19.19 A typical TMDSC scan

the T_g range, as can be seen in the example in Fig. 19.19 which has a particularly narrow glass-transition range and required a $P = 60$ s. The difference between the average specific capacity, as determined from the conventional DSC heat flow, and the complex heat capacity, as determined from TMDSC RHF, in the T_g region is substantial [19.115]. The influence of experimental conditions on TMDSC measurements have been well discussed in the literature [19.120].

The glass transition T_g as observed in DSC scans is well known to depend on thermal history. The classical well-known example is the fact that the T_g value determined by DSC under a heating scan at a given rate r is different from that determined under a cooling scan at a rate $q = r$. Further, T_g determined in a DSC heating scan also depends on aging. The case for As_2Se_3 is well documented [19.60]. Figure 19.20 shows an example in which the T_g values for vitreous As_2Se_3 determined in TMDSC heating and cooling scans are approximately the same for a given modulation

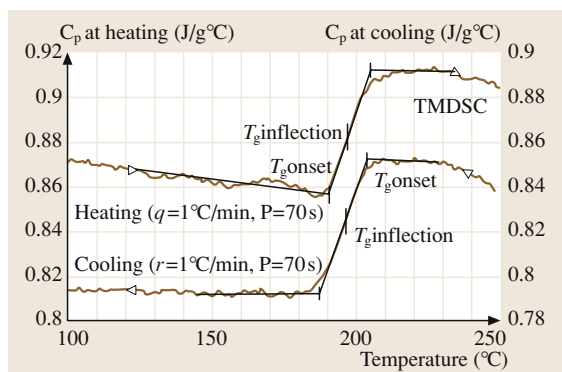


Fig. 19.20 Typical TMDSC heating and cooling scans on As_2Se_3 glass

period P , within experimental errors. T_g depends only on the modulation frequency $\omega (= 2\pi/P)$ provided that the underlying heating or cooling scan is sufficiently slow, usually less than $1^\circ\text{C}/\text{min}$. Other examples are also available in the literature for both glasses and polymers [19.121, 122]. The T_g measurement in the TMDSC experiments represents the relaxation processes associated with the temperature modulation, and not the underlying heating rate as in the conventional DSC. For example, T_g in TMDSC represents a temperature where the relaxation time τ is comparable with the modulation period P . Above T_g , $\tau \ll P$ and below T_g , $\tau \gg P$, and in the T_g region $\tau \approx P$. As P (or ω) is varied, from experiment to experiment, different values of τ satisfy $\tau \approx P$ and hence T_g depends on ω . It is possible to examine the shift in T_g with ω , and then plot $\ln \omega$ versus $1/T_g$ to extract the activation energy of the relaxation time τ , as shown for As_2Se_3 [19.61]. T_g that is observed in conventional DSC experiments is associated with the devitrification (on heating) and vitrification (on cooling) of the structure, whereas T_g in TMDSC experiments represents a dynamic glass-transition temperature; it is important to arrange the experimental conditions so that the two processes are separated in the measurement [19.123].

While a single phenomenological relaxation time provides a convenient description of the overall glass-transformation behavior, a more complete description necessarily involves incorporating a distribution of relaxation times, as discussed in [19.116].

19.5.3 Tzero Technology

It is well known that there is a thermal lag between the sample and temperature-control sensors in the DSC and most DSCs require calibration at the heating rate to be used. Alternatively, calibration at multiple rates is required to correct for the heating rate dependence of the thermal lag. Without this calibration, there is considerable error in the temperature data in changing from one heating rate to another. This problem is especially acute in cooling experiments due to the thermal lag being in the opposite direction to that of heating, thereby doubling the lag effect. The lag problem has been solved in a new DSC called the *Tzero technology*, as developed by TA Instruments [19.124, 125]. With the Q-series DSC system, an additional thermocouple on the DSC sensor allows a complete calibration of the DSC cell and compensation for all thermal lags, including those caused by the encapsulated sample. This means that, for all practical purposes, the DSC can be

calibrated at any reasonable heating rate, and the DSC temperature data will be correct within a few tenths of a Celsius degree for data taken at other heating or cooling rates. Inasmuch as advanced Tzero compensates for the effect of pan thermal mass and coupling, it is possible to calibrate using one pan type and then use another pan type without incurring substantial errors. In summary, since the Tzero technology uses more information in the DSC measurement, it is more accurate under a wider range of conditions compared with the ordinary DSC without calibration under those specific conditions.

The thermal lag error is proportional to heat flow, heating rate, and to the mass of the sample/pan system.

Hence, this error becomes greatest with fast scanning rates, large sample masses, massive sample pans, or sample specimens with an especially high heat capacity. The thermal lag error is also proportional to the thermal resistance between the sample and sensors so it is made worse by using pans made of poor thermal conductivity or pans making poor thermal contact. However, even in ordinary polymer samples, using optimally coupled aluminum pans, the error produced could be more than two Celsius degrees because of poor thermal conductivity; for other samples it could be several times larger. In summary, the Tzero technology has enabled better DSC experiments to be carried out.

References

- 19.1 W. Martienssen, H. Warlimont (eds): *Springer Handbook of Condensed Matter and Materials Data* (Springer, Berlin Heidelberg New York 2005)
- 19.2 O. Madelung: *Semiconductors: Data Handbook*, 3rd edn. (Springer, Berlin Heidelberg New York 2004)
- 19.3 S. Adachi: *Properties of Group IV, III–V and II–VI Semiconductors* (Wiley, Chichester 2005)
- 19.4 P. Debye: *Ann. Phys.* **39**, 789 (1912)
- 19.5 S. O. Kasap: *Principles of Electronic Materials and Devices*, 3rd edn. (McGraw–Hill, Boston 2005)
- 19.6 S. Elliott: *The Physics and Chemistry of Solids* (Wiley, Chichester 1998)
- 19.7 R. B. Stephens: *Phys. Rev. B* **8**, 2896 (1973)
- 19.8 J. De Launay: *Solid State Physics Vol. 2*, ed. by F. Seitz, D. Turnbull (Academic, New York 1956)
- 19.9 K. Ichikawa: *J. Phys. C* **18**, 4631 (1985)
- 19.10 D. W. Van Krevelen, P. J. Hoftyzer: *Properties of Polymer* (Elsevier, Amsterdam 1976)
- 19.11 M. Pyda, E. Nowak-Pyda, J. Mays, B. Wunderlich: *J. Polymer Sci. B* **42**, 4401 (2004)
- 19.12 B. Wunderlich: *Thermochim. Acta* **300**, 43 (1997) and references therein
- 19.13 D. E. Sharp, L. B. Ginther: *J. Am. Ceram. Soc.* **34**, 260 (1951)
- 19.14 S. A. Khalimovskaya-Churkina, A. I. Priven: *Glass Phys. Chem.*, **26**, 531 (2000) and references therein
- 19.15 T. Wagner, S. O. Kasap: *Philos. Mag.* **74**, 667 (1996)
- 19.16 S. Inaba, S. Oda: *J. Non-Cryst. Solids* **325**, 258 (2003)
- 19.17 Y. P. Joshi, G. S. Verma: *Phys. Rev. B* **1**, 750 (1970)
- 19.18 C. J. Glassbrenner, G. Slack: *Phys. Rev.* **134**, A1058 (1964)
- 19.19 M. G. Holland: *The Proceedings of the 7th Int. Conf. Phys. Semicond., Paris* (Dunond, Paris 1964) p.1161. The data were extracted from 19.2 (Fig. 2.11.11) in which the original data were taken from this reference.
- 19.20 C. M. Bhandari, C. M. Rowe: *Thermal Conduction in Semiconductors* (Wiley, New Delhi 1988)
- 19.21 M. P. Zaitlin, A. C. Anderson: *Phys. Rev. Lett.* **33**, 1158 (1974)
- 19.22 C. Kittel: *Phys. Rev.* **75**, 972 (1949)
- 19.23 C. Kittel: *Introduction to Solid State Physics*, 8th edn. (Wiley, New York 2005)
- 19.24 P. B. Allen, J. L. Feldman: *Phys. Rev. Lett.* **62**, 645 (1989)
- 19.25 P. B. Allen, J. L. Feldman: *Phys. Rev. B* **48**, 12581 (1993)
- 19.26 A. Jagannathan, R. Orbach, O. Entin-Wohlman: *Phys. Rev. B* **30**, 13465 (1989)
- 19.27 C. Oligschleger, J. C. Schön: *Phys. Rev. B* **59**, 4125 (1999)
- 19.28 R. C. Zeller, R. O. Pohl: *Phys. Rev. B* **4**, 2029 (1971)
- 19.29 K. Eiermann: *Kolloid Z.* **201**, 3 (1965)
- 19.30 Y. Agari, A. Ueda, Y. Omura, S. Nagai: *Polymer* **38**, 801 (1997)
- 19.31 B. Weidenfeller, M. Höfer, F. Schilling: *Composites A* **33**, 1041 (2002)
- 19.32 B. Weidenfeller, M. Höfer, F. R. Schilling: *Composites A* **35**, 423 (2004)
- 19.33 R. Bube: *Electronic Properties of Crystalline Solids: An Introduction to Fundamentals* (Academic, New York 1974)
- 19.34 A. Jezowski, J. Mucha, G. Pompe: *J. Phys. D* **20**, 1500 (1987)
- 19.35 See Chapter 1 Selected Topic entitled "Thermal Expansion" in the CDROM *Principle of Electronic Materials and Devices*, 3rd Edition, McGraw–Hill, Boston, (2005)
- 19.36 Y. Okada, Y. Tokumaru: *J. Appl. Phys.* **56**, 314 (1984)
- 19.37 J. M. Hutchinson, P. Kumar: *Thermochim. Acta* **391**, 197 (2002)
- 19.38 R. C. Mackenzie: *Thermochim. Acta* **28**, 1 (1979)

- 19.39 W. W. Wedlandt: *Thermal Analysis*, 3 edn. (Wiley, New York 1986) p. 3
- 19.40 B. Wunderlich: *Thermal Analysis* (Academic, New York 1990)
- 19.41 E. F. Palermo, J. Chiu: *Thermochim. Acta* **14**, 1 (1976)
- 19.42 S. Sarig, J. Fuchs: *Thermochim. Acta* **148**, 325 (1989)
- 19.43 W. Y. Lin, K. K. Mishra, E. Mori, K. Rajeshwar: *Anal. Chem.* **62**, 821 (1990)
- 19.44 T. Ozawa: *Thermochim. Acta* **355**, 35 (2000)
- 19.45 J. Wong, C. A. Angell: *Glass, Structure by Spectroscopy* (Marcel Dekker, New York 1976) and references therein
- 19.46 J. Zaryzycki: *Glasses and the Vitreous State* (Cambridge University Press, Cambridge 1991)
- 19.47 J. Jäkle: *Rep. Prog. Phys.* **49**, 171 (1986)
- 19.48 C. A. Angell: *J. Res. Natl. Inst. Stand. Technol.* **102**, 171 (1997)
- 19.49 C. A. Angell, B. E. Richards, V. Velikov: *J. Phys. Cond. Matter* **11**, A75 (1999)
- 19.50 I. Gutzow, B. Petroff: *J. Non-Cryst. Solids* **345**, 528 (2004)
- 19.51 H. N. Ritland: *J. Am. Ceram. Soc.* **37**, 370 (1954)
- 19.52 C. T. Moynihan, A. J. Easteal, M. A. DeBolt, J. Tucker: *J. Am. Cer. Soc.* **59**, 12 (1976)
- 19.53 M. A. DeBolt, A. J. Easteal, P. B. Macedo, C. T. Moynihan: *J. Am. Cer. Soc.* **59**, 16 (1976)
- 19.54 H. Sasabe, C. Moynihan: *J. Polym. Sci.* **16**, 1447 (1978)
- 19.55 O. V. Mazurin: *J. Non-Cryst. Solids* **25**, 131 (1977)
- 19.56 C. T. Moynihan, A. J. Easteal: *J. Am. Ceram. Soc.* **54**, 491 (1971)
- 19.57 H. Sasabe, C. T. Moynihan: *J. Polym. Sci.* **16**, 1447 (1978)
- 19.58 I. Avramov, E. Grantscharova, I. Gutzow: *J. Non-Cryst. Solids* **91**, 386 (1987) and references therein
- 19.59 S. Yannacopoulos, S. O. Kasap: *J. Mater. Res.* **5**, 789 (1990)
- 19.60 S. O. Kasap, S. Yannacopoulos: *Phys. Chem. Glasses* **31**, 71 (1990)
- 19.61 J. Malek: *Thermochim. Acta* **311**, 183 (1998)
- 19.62 S. O. Kasap, D. Tonchev: *J. Mater. Res.* **16**, 2399 (2001)
- 19.63 Z. Cernosek, J. Holubova, E. Cernoskova, M. Liska: *J. Optoelec. Adv. Mater.* **4**, 489 (2002) and references therein
- 19.64 G. Williams, D. C. Watts: *Trans. Faraday Soc.* **66**, 80 (1970)
- 19.65 R. Bohmer, C. A. Angell: *Phys. Rev. B* **48**, 5857 (1993)
- 19.66 W. Kauzmann: *Chem. Rev.* **43**, 219 (1948)
- 19.67 R. F. Boyer: *J. Appl. Phys.* **25**, 825 (1954)
- 19.68 G. M. Bartenev, I. A. Lukianov: *Zh. Fiz. Khim* **29**, 1486 (1955)
- 19.69 T. G. Fox, P. J. Flory: *J. Polym. Sci.* **14**, 315 (1954)
- 19.70 T. G. Fox, S. Loshaek: *J. Polym. Sci.* **15**, 371 (1955)
- 19.71 K. Tanaka: *Solid State Commun.* **54**, 867 (1985)
- 19.72 I. Avramov, T. Vassilev, I. Penkov: *J. Non-Cryst. Solids* **351**, 472 (2005)
- 19.73 I. W. Donald: *J. Non-Cryst. Solids* **345**, 120 (2004)
- 19.74 T. Ozawa: *J. Therm. Anal.* **2**, 301 (1970)
- 19.75 T. Ozawa: *J. Therm. Anal.* **7**, 601 (1975)
- 19.76 T. Ozawa: *J. Therm. Anal.* **9**, 369 (1976)
- 19.77 S. O. Kasap, C. Juhasz: *J. Chem. Soc. Faraday Trans. II* **81**, 811 (1985) and references therein
- 19.78 T. Kemeny, J. Sestak: *Thermochim. Acta* **110**, 113 (1987) and references therein
- 19.79 S. Yannacopoulos, S. O. Kasap, A. Hedayat, A. Verma: *Can. Metall. Q.* **33**, 51 (1994)
- 19.80 A. W. Coats, J. P. Redfern: *Nature* **201**, 68 (1964)
- 19.81 J. Sestak: *Thermochim. Acta* **3**, 150 (1971)
- 19.82 H. E. Kissinger: *J. Res. Natl. Bur. Stand.* **57**, 217 (1956)
- 19.83 H. E. Kissinger: *Anal. Chem.* **29**, 1702 (1957)
- 19.84 S. de la Parra, L. C. Torres-Gonzalez, L. M. Torres-Martínez, E. Sanchez: *J. Non-Cryst. Solids* **329**, 104 (2003)
- 19.85 I. W. Donald, B. L. Metcalfe: *J. Non-Cryst. Solids* **348**, 118 (2004)
- 19.86 W. Luo, Y. Wang, F. Bao, L. Zhou, X. Wang: *J. Non-Cryst. Solids* **347**, 31 (2004)
- 19.87 J. Vazquez, D. Garcia-G. Barreda, P. L. Lopez-Aleman, P. Villares, R. Jimenez-Garay: *J. Non-Cryst. Solids* **345**, 142 (2004) and references therein
- 19.88 A. Pratap, K. N. Lad, T. L. S. Rao, P. Majmudar, N. S. Saxena: *J. Non-Cryst. Solids* **345**, 178 (2004)
- 19.89 K. Matusita, S. Sakka: *Bull. Inst. Chem. Res.* **59**, 159 (1981)
- 19.90 K. Matusita, T. Komatsu, R. Yokota: *J. Mater. Sci.* **19**, 291 (1984)
- 19.91 J. A. Augis, J. W. E. Bennett: *J. Therm. Anal.* **13**, 283 (1978)
- 19.92 H. J. Borchardt, F. Daniels: *J. Am. Ceram. Soc.* **78**, 41 (1957)
- 19.93 G. O. Pilonyan, I. D. Ryabchikov, O. S. Novikova: *Nature* **212**, 1229 (1966)
- 19.94 T. Ozawa: *Polymer* **12**, 150 (1971)
- 19.95 H. S. Chen: *J. Non-Cryst. Solids* **27**, 257 (1978)
- 19.96 M. Reading, D. Elliott, V. L. Hill: *J. Therm. Anal.* **40**, 949 (1993)
- 19.97 M. Reading: *Trends Polym. Sci.* **1**, 248 (1993)
- 19.98 M. Reading, A. Luget, R. Wilson: *Thermochim. Acta* **238**, 295 (1994)
- 19.99 E. Verdonck, K. Schaap, L. C. Thomas: *Int. J. Pharm.* **192**, 3 (1999)
- 19.100 C. M. A. Lopes, M. I. Felisberti: *Polym. Test.* **23**, 637 (2004)
- 19.101 T. Wagner, M. Frumar, S. O. Kasap: *J. Non-Cryst. Solids* **256**, 160 (1999)
- 19.102 P. Boolchand, D. G. Georgiev, M. Micoulaut: *J. Optoelectron. Adv. Mater.* **4**, 823 (2002) and references therein
- 19.103 K. J. Jones, I. Kinshott, M. Reading, A. A. Lacey, C. Nikopoulos, H. M. Pollosk: *Thermochim. Acta* **305**, 187 (1997)
- 19.104 Z. Jiang, C. T. Imrie, J. M. Hutchinson: *Thermochim. Acta* **315**, 1 (1998)
- 19.105 B. Wunderlich: *Thermochim. Acta* **355**, 43 (2000)

- 19.106 H. Huth, M. Beiner, S. Weyer, M. Merzlyakov, C. Schick, E. Donth: *Thermochim. Acta* **377**, 113 (2001)
- 19.107 Z. Jiang, C. T. Imrie, J. M. Hutchinson: *Thermochim. Acta* **387**, 75 (2002)
- 19.108 T. Wagner, S. O. Kasap, K. Maeda: *J. Mater. Res.* **12**, 1892 (1997)
- 19.109 I. Okazaki, B. Wunderlich: *J. Polym. Sci.* **34**, 2941 (1996)
- 19.110 L. Thomas, A. Boller, I. Okazaki, B. Wunderlich: *Thermochim. Acta* **291**, 85 (1997)
- 19.111 L. Thomas: NATAS Notes (North American Thermal Analysis Society, Sacramento, CA, USA) **26**, 48 (1995)
- 19.112 B. Hassel: NATAS Notes (North American Thermal Analysis Society, Sacramento, CA, USA) **26**, 54 (1995)
- 19.113 J. M. Hutchinson, S. Montserrat: *J. Therm. Anal.* **47**, 103 (1996)
- 19.114 J. M. Hutchinson, S. Montserrat: *Thermochim. Acta* **305**, 257 (1997)
- 19.115 J. M. Hutchinson: *Thermochim. Acta* **324**, 165 (1998)
- 19.116 J. M. Hutchinson, S. Montserrat: *J. Therm. Anal.* **377**, 63 (2001) and references therein
- 19.117 A. Boller, C. Schick, B. Wunderlich: *Thermochim. Acta* **266**, 97 (1995)
- 19.118 J. M. Hutchinson, A. B. Tong, Z. Jiang: *Thermochim. Acta* **335**, 27 (1999)
- 19.119 D. Tonchev, S. O. Kasap: *Mater. Sci. Eng.* **A328**, 62 (2002)
- 19.120 J. E. K. Schawe: *Thermochim. Acta* **271**, 127 (1996)
- 19.121 P. Kamasa, M. Pyda, A. Buzin, B. Wunderlich: *Thermochim. Acta* **396**, 109 (2003)
- 19.122 D. Tonchev, S. O. Kasap: Thermal Characterization of Glasses and Polymers by Temperature Modulated Differential Scanning Calorimetry: Glass Transition Temperature. In: *High Performance Structures and Materials II*, ed. by C. A. Brebbia, W. P. De Wilde (WIT, Southampton, UK 2004) pp. 223–232
- 19.123 S. Weyer, M. Merzlyakov, C. Schick: *Thermochim. Acta* **377**, 85 (2001)
- 19.124 L. E. Waguespack, R. L. Blaine: Design of a new DSC cell with Tzero technology. In: *Proceedings of the 29th North American Thermal Analysis Society, St. Louis, September 24–26*, ed. by K. J. Kociba (NATAS, Sacramento 2001) pp. 722–727
- 19.125 R. L. Danley: *Thermochim. Acta* **395**, 201 (2003)

20. Electrical Characterization of Semiconductor Materials and Devices

Semiconductor materials and devices continue to occupy a preeminent technological position due to their importance when building integrated electronic systems used in a wide range of applications from computers, cell-phones, personal digital assistants, digital cameras and electronic entertainment systems, to electronic instrumentation for medical diagnostics and environmental monitoring. Key ingredients of this technological dominance have been the rapid advances made in the quality and processing of materials – semiconductors, conductors and dielectrics – which have given metal oxide semiconductor device technology its important characteristics of negligible standby power dissipation, good input–output isolation, surface potential control and reliable operation. However, when assessing material quality and device reliability, it is important to have fast, nondestructive, accurate and easy-to-use electrical characterization techniques available, so that important parameters such as carrier doping density, type and mobility of carriers, interface quality, oxide trap density, semiconductor bulk defect density, contact and other parasitic resistances and oxide electrical integrity can be determined. This chapter describes some of the more widely employed and popular techniques that are used to determine these important parameters. The techniques presented in this chapter range in both complexity and test structure requirements from simple current–voltage measurements to more sophisticated

20.1 Resistivity	410
20.1.1 Bulk Resistivity	410
20.1.2 Contact Resistivity	415
20.2 Hall Effect	418
20.2.1 Physical Principles	419
20.2.2 Hall Scattering Factor	420
20.3 Capacitance–Voltage Measurements	421
20.3.1 Average Doping Density by Maximum–Minimum High- Frequency Capacitance Method ...	421
20.3.2 Doping Profile by High-Frequency and High–Low Frequency Capacitance Methods..	422
20.3.3 Density of Interface States	424
20.4 Current–Voltage Measurements	426
20.4.1 I – V Measurements on a Simple Diode	426
20.4.2 I – V Measurements on a Simple MOSFET	426
20.4.3 Floating Gate Measurements	427
20.5 Charge Pumping	428
20.6 Low-Frequency Noise	430
20.6.1 Introduction	430
20.6.2 Noise from the Interfacial Oxide Layer ..	431
20.6.3 Impedance Considerations During Noise Measurement	432
20.7 Deep-Level Transient Spectroscopy	434
References	436

low-frequency noise, charge pumping and deep-level transient spectroscopy techniques.

The continued evolution of semiconductor devices to smaller dimensions in order to improve performance – speed, functionality, integration density and reduced cost – requires layers or films of semiconductors, insulators and metals with increasingly high quality that are well-characterized and that can be deposited and patterned to very high precision. However, it is not always the case that improvements in the quality of ma-

terials have kept pace with the evolution of integrated circuit down-scaling. An important aspect of assessing the material quality and device reliability is the development and use of fast, nondestructive and accurate electrical characterization techniques to determine important parameters such as carrier doping density, type and mobility of carriers, interface quality, oxide trap density, semiconductor bulk defect density, con-

tact and other parasitic resistances and oxide electrical integrity. This chapter will discuss several techniques that are used to determine these important parameters. However, it is not an extensive compilation of the electrical techniques currently used by the research and development community; rather, it presents a discussion of some of the more widely used and popular ones [20.1–4].

An important aspect of electrical characterization is the availability of appropriate test components [20.1–4]. In this chapter, we concentrate on discussing techniques that use standard test devices and structures. In addition, we will use the MOSFET whenever possible because they are widely available on test chips. This is also motivated by the fact that MOSFETs continue to dominate the semiconductor industry for a wide range of applications from memories and microprocessors to signal and imaging processing systems [20.5]. A key reason for this dominance is the excellent quality of the silicon wafers and the silicon–silicon dioxide interface, both of which play critical roles in the performance and reliability of the device. For example, if the interface has many defects or interface states, or it is rough, then the device’s carrier mobility decreases, low-frequency noise increases and its performance and reliability degrades. In particular, it is not only the interface that is important, but also the quality of the oxide; good-quality oxide prevents currents from flowing between the gate and substrate electrodes through the gate oxide. Both interface and oxide quality allows for excellent isolation between the input and output terminals of the MOSFETs, causing it to behave as an almost ideal switch. Therefore, it is important to have good experimental tools to study the interface properties and the quality of the gate dielectric.

Electrical characterization of semiconductors and the semiconductor–dielectric interface is important for a variety of reasons. For example, the defects at and in the interfacial oxide layer in silicon–silicon dioxide (Si–SiO₂) systems and in the bulk semiconductor play critical roles in their low-frequency noise, independent of whether the device is surface-controlled such as a MOSFET, or a bulk transport device such as a polysilicon emitter bipolar junction transistor (PE BJT). These defects can affect the charge transfer efficiency in charge coupled devices (CCDs), p–n photodiodes or CMOS imagers, and can be the initiation point of catastrophic failure of oxides. Interface and bulk states can act as scattering centers to reduce the mobility in MOSFETs, thus affecting their performance parameters such as switching speed, transconductance and noise.

This chapter is devoted to the electrical characterization of semiconductors, insulators and interfaces. In the first part (Sects. 20.1 and 20.2), the basic electrical properties of materials (such as resistivity, concentration and mobility of carriers) are studied. The main measurement techniques used to determine these electrical parameters are presented. Due to its increasing importance in modern ultrasmall geometry devices, electrical contacts are also studied. All of the characterization techniques presented in this first part are associated with specially designed test structures. In the second part (Sects. 20.4 to 20.7), we use active components such as capacitors, diodes and transistors (mainly MOSFETs) in order to determine more specific electrical parameters such as traps, oxide quality and noise level that are associated with material or devices. Of course this involves specific measurement techniques that are often more sophisticated than those discussed in the previous two sections.

20.1 Resistivity

Resistivity is one of the most important electrical parameters of semiconductors [20.1–4]. First, we present the basic physical relations concerning the bulk resistivity. The main electrical measurement techniques are then described: the two oldest ones that are still relevant today – the *four-point-probe* technique and the *van der Pauw* technique – and then the *spreading resistance* technique. Second, because it is closely linked with bulk resistivity measurement techniques and it is increasingly important in modern ultrasmall geometry devices, contact resistivity will be presented. Special attention will be given to

Kelvin contact resistance (KCR) measurement and the *transmission line measurement (TLM)* techniques.

20.1.1 Bulk Resistivity

Physical Approach, Background and Basics
The bulk resistivity ρ is an intrinsic electrical property related to carrier drift in materials such as metals and semiconductors [20.6]. From a macroscopic point of view, the resistivity ρ can be viewed as the normalization of the bulk resistance (R) by its geometrical dimensions

– the cross-sectional area ($A = Wt$) through which the current flows, and the distance between the two ideal contacts L , as shown in Fig. 20.1. The resistivity is given by

$$\rho = \frac{RA}{L} \text{ in } \Omega\text{m or commonly } \Omega\text{cm.} \quad (20.1)$$

For thin semiconductor layers, the sheet resistivity ρ_s is often used instead of the bulk resistivity ρ . The sheet resistivity ρ_s is the bulk resistivity divided by the sample's thickness t . This normalized parameter is related to the resistance of a square of side L . For this particular geometry in Fig. 20.1, since $A = Wt$, then $\rho_s = R_{\square}$, the sheet resistance. The unit of sheet resistance is Ω/square or Ω/\square . The parameter R_{\square} is convenient for integrated circuit designers because it allows them to quickly design the geometry for a specific value of resistance using very thin implanted or diffused semiconductor regions or polycrystalline layers. Resistivity (or its inverse, the conductivity σ in $\Omega^{-1}\text{cm}^{-1}$ or S/cm) and its variation with temperature is often used to classify material into metals, semiconductors and insulators.

Since different semiconductors can have the same resistivity, and also different values of resistivity can be found for a given semiconductor, depending on how it is processed for example, then resistivity is not a fundamental material parameter. From solid state theory, in the case of homogeneous semiconductor materials, the resistivity expresses the proportionality between the applied electric field E and the drift current density J ; that is, $J = (1/\rho) E$. It can be defined by the microscopic relation:

$$\rho = \frac{1}{q(n\mu_n + p\mu_p)}, \quad (20.2)$$

where q is the electronic charge, n and p are the free electron and hole concentrations, and μ_n and μ_p are the electron and hole drift mobilities, respectively. In this way, the resistivity is related to fundamental semiconductor parameters: the number of free carriers, and their

ability to move in the lattice when an electric field is applied.

In n-type or donor (N_D)-doped, or p-type or acceptor (N_A)-doped semiconductors, the free carrier densities are determined by the ionized impurities (N_D or $N_A \gg$ the intrinsic carrier concentration n_i), then (20.2) can be simplified to:

$$\rho \approx \frac{1}{qn\mu_n}, \text{ for an n-type semiconductor,} \quad (20.3a)$$

$$\text{and } \rho \approx \frac{1}{qp\mu_p}, \text{ for an p-type semiconductor.} \quad (20.3b)$$

In the following sections, only single-type semiconductors will be studied. This corresponds to most semiconductor materials used in electronic and optoelectronic devices because either $N_D \gg N_A$ or $N_A \gg N_D$ in a typical semiconductor layer.

Measurement Techniques

The simplest way to determine bulk resistivity is to measure the voltage drop along a uniform semiconductor bar through which a DC current I flows, as shown in Fig. 20.1. Thus, the measured resistance and knowledge of the geometrical dimensions can lead to an estimate for the bulk resistivity according to (20.1). Unfortunately the measured resistance (R_{mea}) includes the unexpected contact resistance ($2R_c$), which can be significant for small-geometry samples because R_c is strongly dependent on the metal-semiconductor structure. Therefore, special processing technologies are used to minimize the influence of R_c (see Sect. 20.1.2). Now, the measured resistance is expressed as

$$R_{\text{mea}} = R + 2R_c. \quad (20.4)$$

If probes are used instead of large metal-semiconductor contacts, then the spreading resistance (R_{sp}) under the two probes must also be added, as shown in Fig. 20.2. In this case, (20.4) becomes

$$R_{\text{mea}} = R + 2R_{\text{sp}} + 2R_c, \quad (20.5)$$

where R_{sp} for a cylindrical contact of radius r , and for a semi-infinite sample, it can be expressed by

$$R_{\text{sp}} = \frac{\rho}{4r}. \quad (20.6)$$

For a hemispherical contact of radius r , R_{sp} is given by

$$R_{\text{sp}} = \frac{\rho}{2\pi r}. \quad (20.7)$$

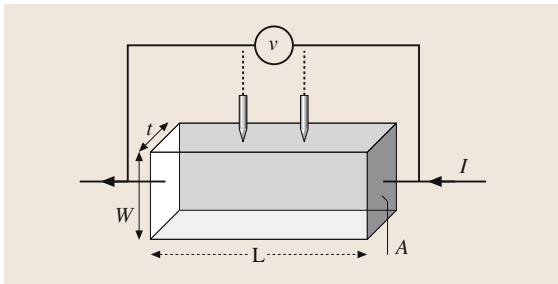


Fig. 20.1 Bulk resistance and its geometrical dimensions

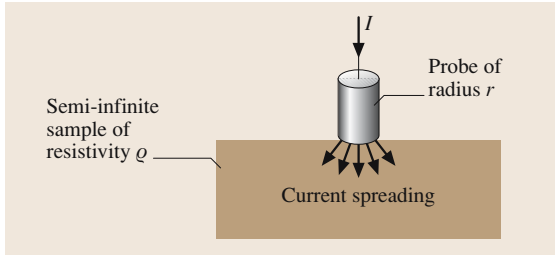


Fig. 20.2 Spreading resistance associated with a probe contact

In both cases, it is very difficult to provide a direct measurement or an accurate model of the contact resistance. So determining the bulk resistivity by this approach is not recommended, except when the spreading resistance is the dominant term in (20.5) and when (20.6) or (20.7) can be applied. In this case, the resistivity is determined by the *spreading resistance* technique measurement. Nevertheless, despite the lack of accuracy of the two contact techniques, it can be sufficient for monitoring some process steps and it is often used in the semiconductor industry as a process monitor.

Four-Point Probe Technique. In order to eliminate or at least minimize the contact contribution to the measured resistance value, techniques based on separate current injection and voltage drop measurements have been developed. First, the two-probe technique can be used, as reported in Fig. 20.1. This measurement is very simple, but it is affected by several parameters: lateral contact geometry, probe spacing, and minority carrier injection near the lateral contacts. The main disadvantage of this technique is the need for lateral contacts. This requirement is overcome with the four-point probe technique, where two probes are used for current injection and the other two probes are used to measure the voltage drop. The more usual probe geometry configuration is when the four probes are placed in a line, as shown in Fig. 20.3.

The voltage at probe 2, V_2 , induced by the current flowing from probe 1 to probe 4 is given by:

$$V_2 = \frac{\rho I}{2\pi} \cdot \left(\frac{1}{s_1} - \frac{1}{s_2 + s_3} \right). \quad (20.8a)$$

The voltage at probe 3 is:

$$V_3 = \frac{\rho I}{2\pi} \cdot \left(\frac{1}{s_1 + s_2} - \frac{1}{s_3} \right). \quad (20.8b)$$

Then, by measuring $V = V_2 - V_3$, the voltage drop between probes 2 and 3, and the current I through probes 1 and 4, the resistivity can be determined using (20.8a)

and (20.8b) as

$$\rho = \frac{2\pi V/I}{\left(\frac{1}{s_1} + \frac{1}{s_2} - \frac{1}{s_2 + s_3} - \frac{1}{s_1 + s_2} \right)} \quad (20.9)$$

Thus, a direct measurement of the resistivity can be made using a high-impedance voltmeter and a current source. When the probe spacings are equal ($s_1 = s_2 = s_3 = s$), which is the most practical case, then (20.9) becomes

$$\rho = 2\pi s \cdot \frac{V}{I}. \quad (20.10)$$

Equations (20.9) and (20.10) are valid only for semi-infinite samples; that is, when both t and the sample surface are very large ($\rightarrow \infty$), and the probes' locations must be far from any boundary. Because these relations can be applied only to large ingots, then in many cases a correction factor f must be introduced in order to take into account the finite thickness and surface of the sample and its boundary effects. Further, for epitaxial layers, f must also consider the nature of the substrate – whether it is a conductor or an insulator. Thus, (20.10) becomes

$$\rho = 2\pi s \cdot \frac{V}{I} \cdot f. \quad (20.11)$$

For a thin semiconductor wafer or thin semiconducting layer deposited on an insulating substrate, and for the condition $t < s/2$, which represents most practical cases because the probe spacing s is usually on the order of a millimeter, then the correction factor due to the thickness is

$$f = \frac{(t/s)}{2 \ln 2} \text{ so that } \rho = 4.532t \frac{V}{I}. \quad (20.12)$$

The noninfinite sample surface must be corrected if the ratio of the wafer diameter to the probe spacing is not

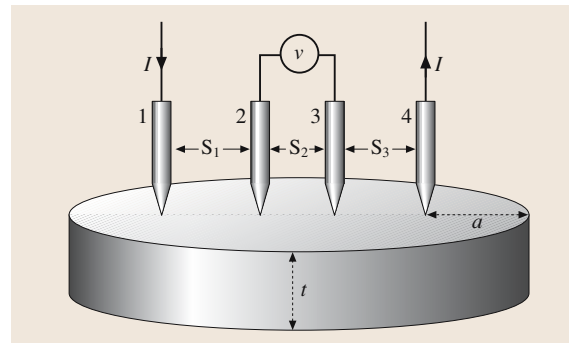


Fig. 20.3 Linear four-point probe configuration. The sample thickness is t and a is the distance from the edge or boundary of the sample

greater than 40, otherwise a correction factor of less than unity has to be introduced [20.3].

If the probe header is too close to any boundary, then (20.12) is no longer valid and another correction factor must be introduced. This correction factor is close to 1 until the ratio a/s is greater than 2, where a is the distance from the edge of the sample that is shown in Fig. 20.3. A study of various 8 inch n- and p-type silicon wafers have demonstrated that the edge exclusion limit is 5 mm [20.7].

In the case of a different arrangement of probes, for instance a square array or when a different measurement configuration of the four-point collinear probes is used, such as current injection between probes 1 and 3, other specific correction factors are required. Here, rather than detail all the different correction factors, complementary information can be found in Chap. 4 of [20.4] and Chap. 1 of [20.3].

Taking into account the appropriate correction factors as well as some specific material parameters such as hardness or surface oxidation, it is possible to map the resistivities of different types of semiconductor wafers or deposited semiconductor layers with an accuracy better than 1% over a large range of resistivity values using commercial equipment and appropriate computational techniques.

Van der Pauw Technique. Based on the same basic principle of separating the current injection and voltage measurement, the *van der Pauw* [20.8] measurement technique allows for the determination of resistivity on a sample of arbitrary shape using four small contacts placed on the periphery, as shown in Fig. 20.4. Then, the resistivity of a uniform sample of thickness t is given

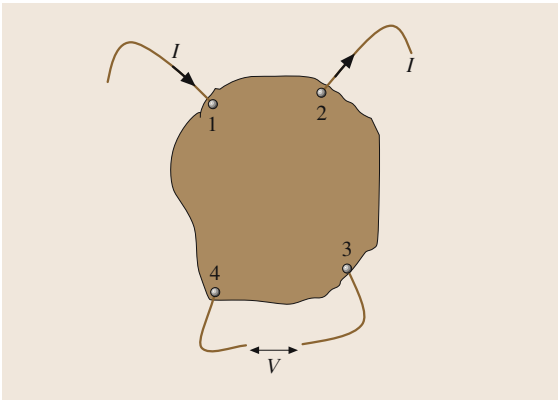


Fig. 20.4 van der Pauw method for an arbitrarily shaped sample

by

$$\rho = \frac{\pi t}{\ln 2} \frac{(R_A + R_B)}{2} f. \quad (20.13)$$

Here, R_A and R_B are resistances measured by injecting current on two adjacent contacts and by measuring the voltage drop on the two remaining ones. With the notation in Fig. 20.4, one can define

$$R_A = \frac{V_3 - V_4}{I_{1,2}}, \quad R_B = \frac{V_4 - V_1}{I_{2,3}}, \quad (20.14)$$

f is a correction factor that is

$$\text{a function of the ratio} \quad R_f = \frac{R_A}{R_B}.$$

with R_f obtained from

$$\frac{(R_f - 1)}{(R_f + 1)} = \frac{f}{\ln 2} \cdot \operatorname{arccosh} \left(\frac{\exp(\ln 2 / f)}{2} \right). \quad (20.15)$$

In the case of samples with symmetrical geometries, and when the contacts are also symmetrical, as shown in Fig. 20.5, then $R_A = R_B$, $R_f = 1$ and $f = 1$, and (20.13) becomes

$$\rho = \frac{\pi t}{\ln 2} R_A = 4.532tR_A \quad (20.16)$$

In order to minimize errors caused by the finite dimensions of the contacts (since ideally the contact area should be zero) and the finite thickness of the sample, then the distance between the contacts must be larger than both the diameter and the thickness of the contact. Also, the cloverleaf configuration in Fig. 20.5d is recommended to prevent contact misalignment, but this configuration requires a more complicated patterning technology.

The main advantage of the *van der Pauw* technique compared to the four-point probe technique is its use of a smaller area for the test structure. Therefore, this measurement technique is often used in integrated circuit technology. Also, because of its simple structure, the Greek cross configuration in Fig. 20.5b is

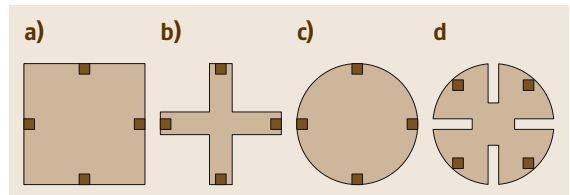


Fig. 20.5a–d Symmetrical van der Pauw structures: (a) square, (b) Greek cross, (c) circle and (d) cloverleaf

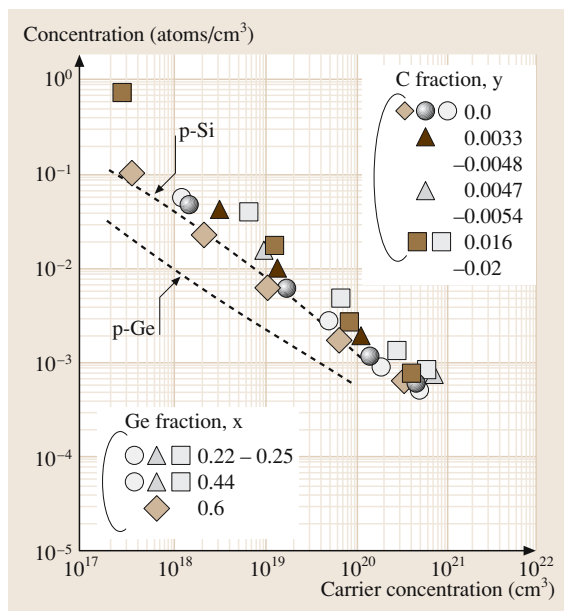


Fig. 20.6 Resistivity versus carrier concentration in $\text{Si}_{1-x}\text{Ge}_x\text{C}_y$ films obtained using a cloverleaf van der Pauw structure. These results are from [20.10]

widely used (experimental results obtained on SiGeC epitaxial layers are reported in Fig. 20.6 as an example). However, when narrow arms are used, current crowding at the corners may have a significant influence, and in this case a different Greek cross layout

can be considered to reduce this current crowding effect [20.9].

Spreading Resistance Technique. The spreading resistance technique is based on the modeling of current spreading from a probe tip or a small metallic contact and flowing into a bulk semiconductor, as shown previously in Fig. 20.2. Equations (20.6) and (20.7) presented above are for cylindrical probes and hemispherical probes, respectively. Basically, the principle of this method is opposite to the previous four-contact techniques where the separation of the current injection from the measured voltage drop was used to avoid the spreading resistance. Here, the spreading resistance is expected to be the dominant term in (20.5). Only two contacts are needed: two closely aligned probes, a small top contact probe or a metallic contact and a large bottom contact. In the first case, surface mapping can be performed, but the main use of this compact probe configuration is for resistivity profiling using a bevelled sample [20.3]. The second configuration has been used to measure the substrate resistivity of silicon integrated circuits where simple test structures – for example the square top contact of $25\ \mu\text{m} \times 25\ \mu\text{m}$ and $50\ \mu\text{m} \times 50\ \mu\text{m}$ shown in [20.12] – have been included on a test chip.

More recently, semiconductor resistivity has been nanocharacterized using scanning spreading resistance microscopy (SSRM) with a standard atomic force microscope (AFM) of lateral resolution of 10 to 20 nm. A SSRM image of a $0.5\ \mu\text{m}$ nMOSFET is given in

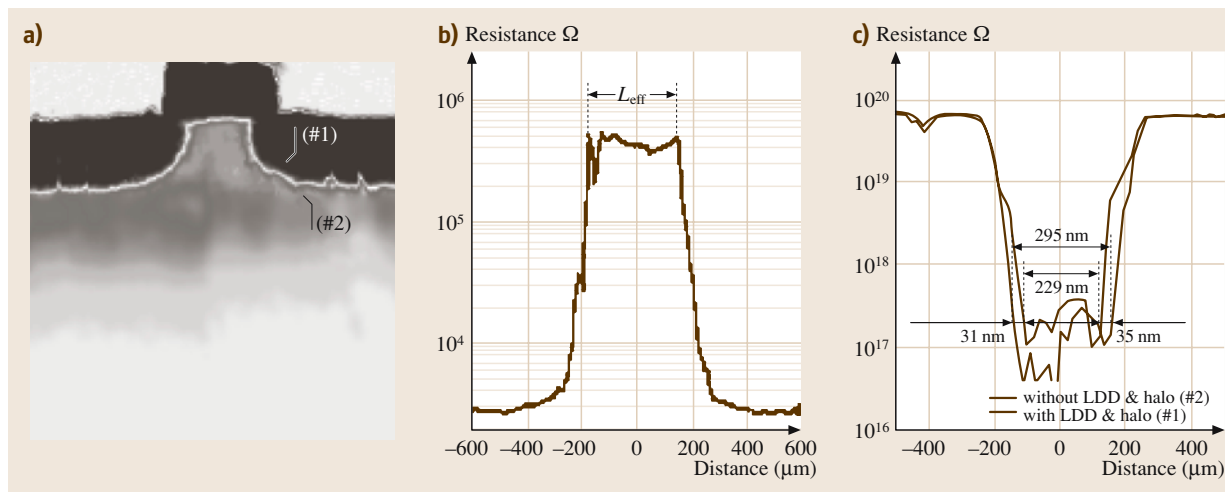


Fig. 20.7 (a) SSRM resistance image (scan size: $1.5\ \mu\text{m} \times 1.5\ \mu\text{m}$) of a $0.5\ \mu\text{m}$ nMOSFET; (b) lateral section taken 10 nm under the gate oxide of the same transistor; (c) lateral carrier concentration profiles measured with SSRM 10 nm under the gate oxide for two $0.35\ \mu\text{m}$ nMOSFETs (with and without halo and LDD process). After [20.11]

Fig. 20.7a [20.11]. The resistance is low in the highly doped regions (dark) and high in the lower doped regions (bright): source, drain, gate and well regions are clearly observed in Fig. 20.7a or the resistance profile in Fig. 20.7b. With such a high resolution, scanning the lateral and vertical diffusion of dopants in active regions of submicron transistors is possible. An example is shown in Fig. 20.7c where the extra implantations [halo and lightly doped drain (LDD)] in a $0.35\text{ }\mu\text{m}$ nMOSFET process are clearly visible and result in a change of L_{eff} from 295 nm without extra implantations to 229 nm with the extra implantations.

20.1.2 Contact Resistivity

The contact resistance of an active device and interconnection becomes larger as the dimensions are scaled down. As a consequence, the performance of single transistors as well as integrated circuits can be seriously limited by increasing RC time constants and power consumption. This is of major interest for the semiconductor industry, as reported by the International Technology Roadmap for Semiconductors, ITRS 2001 [20.5], and in [20.13].

Contact Resistance Elements

Basically, the contact resistance R_c is the resistance localized from a contact pad, a probe or from the metalization process to an active region. However, it does not include all of the access resistances between these two regions, as shown in Fig. 20.8a for a horizontal contact and Fig. 20.8b for a vertical contact.

Starting from the contact pad (Fig. 20.8a and Fig. 20.9), the contact resistance includes the resistance of the metal R_m , the interfacial metal-semiconductor resistance R_i , and the resistance associated with the semiconductor just below the contact in the contact region R_{sc} . Thus, the contact resistance can be expressed

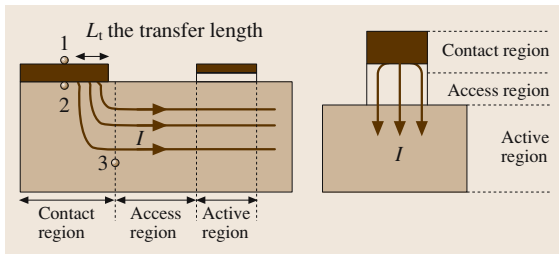


Fig. 20.8 (a) Horizontal contact and (b) vertical contact. Black indicates the metallic conductor, white the semiconductor material or an insulator

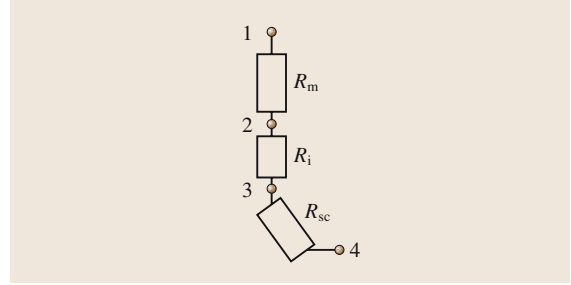


Fig. 20.9 Different components of the contact resistance

as

$$R_c = R_m + R_i + R_{sc} . \quad (20.17)$$

The last component R_{sc} cannot be accurately defined because the boundary between the contact and access regions is very difficult to determine due to (for example) interdiffusion of metal and semiconductor atoms, and because the current flow into this region is not homogeneous due to current spreading and lateral or vertical current crowding at the periphery of the contact. The relative importance of each component of R_c is strongly dependent on different parameters of the process itself – annealing temperature, doping density and the geometry used (lateral or vertical).

When comparing different contact technologies and different contact areas, the most convenient parameter to use is the contact resistivity ρ_c , which is referred to as the specific contact resistance in $\Omega\text{ cm}^2$, and ρ_c is given by

$$\rho_c = R_c A_{\text{ceff}} , \quad (20.18)$$

where A_{ceff} is the effective contact area; that is, the current injection area. The concept of an effective contact area can be approximated by the contact geometry in the case of a vertical contact in Fig. 20.8b. However, A_{ceff} is more difficult to specify for a lateral contact, where a transfer length L_T , representing the length where the current flow transfers from the contact into the semiconductor just underneath, must be introduced, as shown in Fig. 20.8a. L_T is defined as the length over which the voltage drops to e^{-1} of its value at the beginning of the contact [20.3], and is given by

$$L_T = \sqrt{\frac{\rho_c}{\rho_{sc}}} , \quad (20.19)$$

where ρ_{sc} is the sheet resistivity of the semiconductor below the contact.

Because of its various components, it is difficult to accurately model the contact resistivity. Nevertheless,

a theoretical approach to the interfacial resistivity [see R_i in (20.17)], ρ_i , can be determined from the well-known Schottky theory of metal–semiconductor contacts. The interfacial resistivity ρ_i is defined by

$$\rho_i = \left. \frac{\partial V}{\partial J} \right|_{V=0}. \quad (20.20)$$

This metal–semiconductor structure is equivalent to an abrupt p-n junction. According to the Schottky theory (for more details see Chap. 5 of [20.14]), the J – V characteristic of a metal–semiconductor contact in the case of a low-doped semiconductor is given by

$$J = A^* T^2 \exp\left(-\frac{q\phi_B}{kT}\right) \left[\exp\left(\frac{qV}{kT}\right) - 1 \right], \quad (20.21)$$

where A^* is Richardson's constant, and T the absolute temperature. ϕ_B is the barrier height formed at the metal–semiconductor interface – the difference between the vacuum level and the Fermi level of the metal and of the semiconductor materials respectively, and ϕ_B is given by

$$\phi_B = \phi_M - \chi \quad (20.22)$$

where ϕ_M is the metal work function and χ the semiconductor electron affinity.

The energy band diagram for a low-doped n-type semiconductor–metal contact is shown in Fig. 20.10. In this case, the current transport is dominated by the thermionic emission current, resulting in a rectifying contact.

Thus, when the conduction mechanism is controlled by the thermionic emission, the interfacial resistivity in (20.20) is simply obtained from the derivative of (20.21), and $\rho_{i,TE}$ is

$$\rho_{i,TE} = \frac{k}{qA^*T} \exp\left(\frac{q\phi_B}{kT}\right) \quad (20.23)$$

Due to the presence of surface states, the barrier height ϕ_B is positive and weakly dependent on the metal–semiconductor material. ϕ_B is $\approx 2E_g/3$ for an n-type

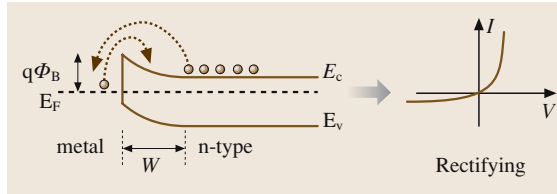


Fig. 20.10 Energy-band diagram of an n-type semiconductor–metal contact and related rectifying contact. W is the width of the depletion layer

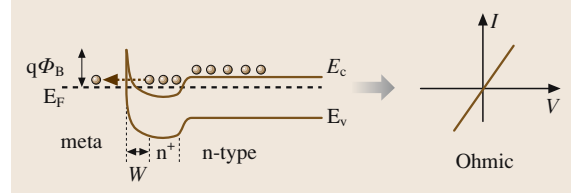


Fig. 20.11 Energy-band diagram of an n^+ -n semiconductor–metal structure and related ohmic contact

semiconductor and $\approx E_g/3$ for a p-type semiconductor. Therefore, high values of interfacial resistivity $\rho_{i,TE}$ are usually obtained except when narrow bandgap semiconductors are used.

The way to fabricate ohmic contacts with low contact resistivity values is to process the metal on a heavily doped semiconductor layer. In this case, the depletion width decreases ($W \approx N_D^{-1/2}$) and the probability of carrier tunneling through the barrier increases. Thus, the conduction mechanism is dominated by tunneling, as shown in Fig. 20.11.

The electron tunneling current is expressed as

$$J_{\text{tun}} \approx \exp\left(\frac{q\phi_B}{E_{00}}\right), \quad (20.24)$$

where

$$E_{00} = \frac{q\hbar}{2} \sqrt{\frac{N_D}{\epsilon_s m_n^*}}, \quad (20.25)$$

ϵ_s is the permittivity of the semiconductor and m_n^* is the effective mass of the electron.

From (20.20), (20.24) and (20.25), the interfacial resistivity $\rho_{i,T}$ is found to be

$$\rho_{i,T} \propto \frac{2\sqrt{\epsilon_s m_n^*}}{\hbar} \cdot \frac{\phi_B}{\sqrt{N_D}}. \quad (20.26)$$

Comparing $\rho_{i,TE}$ from (20.23) to $\rho_{i,T}$ from (20.26), we see that a highly doped layer can significantly reduce the interfacial resistivity. For $N_D \geq 10^{19} \text{ cm}^{-3}$, the tunneling process dominates the interfacial resistivity, while for $N_D \leq 10^{17} \text{ cm}^{-3}$, the thermionic emission current is dominant.

As most semiconductors such as Si, SiGe, GaAs, InP are of relatively wide bandgap, the deposition of a heavily doped layer before the metallization is commonly used in order to form a tunneling contact. For compound semiconductor manufacturing processes, the contact layer is generally formed from the same semiconductor material, or at least from the same material as the substrate. For silicon and related materials such

as SiGe alloys or polysilicon, silicidation techniques are commonly used to make the contact layer with very thin silicide layers such as CoSi₂ or TiSi₂ layers.

Measurement Techniques

As mentioned above, it is difficult to accurately model the contact resistance, so direct measurements of the contact resistance or of the contact resistivity are of great importance. The two main test structures used to

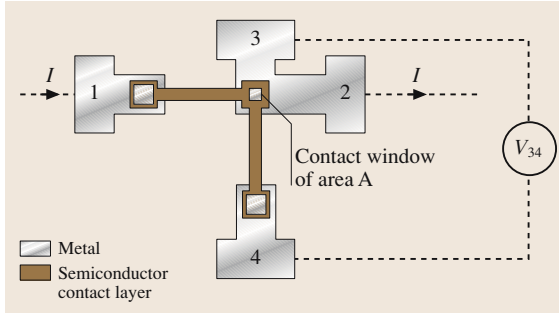


Fig. 20.12 Cross Kelvin resistor test structure

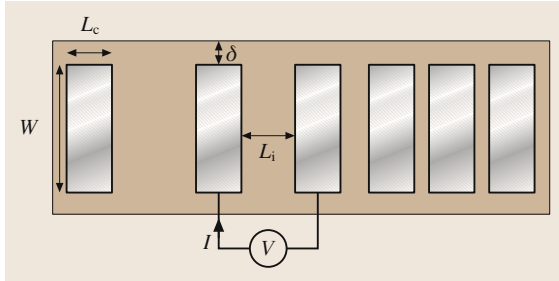


Fig. 20.13 Transmission line model (TLM) test structure

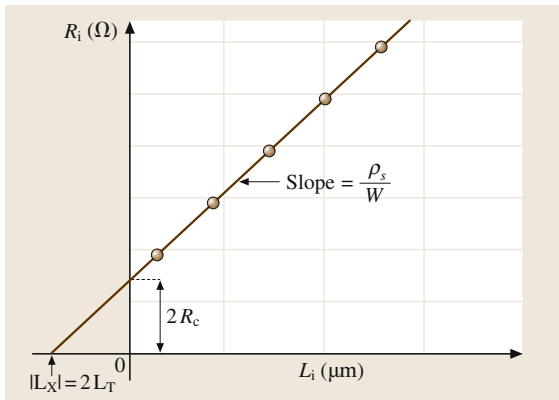


Fig. 20.14 Determination of the sheet resistivity and characterization of the contact using a TLM test structure

determine contact characteristics will now be discussed: the *cross Kelvin resistor (CKR) test structure* and the *transmission line model (TLM) structure*.

Kelvin Test Structure. The Kelvin test structure, also referred to as the cross Kelvin resistor (CKR) test structure, is shown in Fig. 20.12. The contact resistance R_c is determined from the potential drop in the contact window (V_{34}) when a current I is forced through the contact window from contact pad 1 to pad 2, and R_c is

$$R_c = \frac{V_{34}}{I} . \quad (20.27)$$

Therefore, a measure of R_c and knowledge of the contact area A allows for direct extraction of the contact resistivity ρ_c , given by

$$\rho_c = R_c A . \quad (20.28)$$

This basic approach is not valid when parasitic effects are present. One of the main problems is current crowding around the contact. In order to extract accurate values for the contact resistivity using Kelvin test structures, it is necessary to take into account the two-dimensional current-crowding effect. This is achieved using the results from numerical simulations [20.15]. Nevertheless, the development of ohmic contacts with very low values of contact resistivities require complex technology with different materials and usually with several interfaces. In this case, a large discrepancy between the extracted and the measured contact parameters can be found [20.15, 16]. To improve the accuracy, three-dimensional models are now used to take into account the different interfacial and vertical parasitic effects [20.17].

Transmission Line Model Test Structures. The transmission line model test structure (TLM) consists of depositing a metal grid pattern of unequal spacing L_i between the contacts. This leads to a scaled planar resistor structure. Each resistor changes only by its distance L_i between two adjacent contacts, as shown in Fig. 20.13, and it can be expressed by

$$R_i = \frac{\rho_s L_i}{W} + 2R_c \quad (20.29)$$

Then, by plotting the measured resistances as a function of the contact spacing L_i , and according to (20.29), the layer sheet resistivity ρ_s and the contact resistance R_c can be deduced from the slope and from the intercept at

$L_i = 0$ respectively, as shown in Fig. 20.14:

$$\begin{aligned} \text{Slope} &= \frac{\rho_s}{W}; R_i(\text{intercept}) = 2R_c; \\ \text{and } |L_i(\text{intercept})| &= 2L_T \end{aligned} \quad (20.30)$$

As discussed in Sect. 20.1.2, the most suitable parameter for characterizing a contact is its contact resistivity (ρ_c) or the specific contact resistance ($R_c A_{\text{ceff}}$), given by

$$\rho_c = R_c A_{\text{ceff}} = R_c W L_T \quad (20.31)$$

As shown in Fig. 20.8a, for a planar resistor, the effective contact area requires the notion of the transfer length L_T . According to (20.19), and assuming that the sheet resistance under the contact ρ_{sc} is equal to the

sheet resistance between the contacts ρ_s , then L_T can be expressed by (20.19).

Therefore, the substitution of R_c into (20.31) in (20.29) leads to

$$R_i = \frac{\rho_s}{W} L_i + \frac{\rho_s}{W} 2L_T \quad (20.32)$$

Now, extrapolation to $R_i = 0$ allows us to determine the value of L_T . The main advantage of the *TLM* method is its ability to give two main electrical parameters, the resistivity of the semiconductor contact layer ρ_s and the contact resistance R_c . However, this is done at the expense of a questionable assumption that the sheet resistance under the contact must be equal to the sheet resistance between the contacts. More on this technique can be found in [20.3].

20.2 Hall Effect

As mentioned before, the resistivity of a semiconductor is not a fundamental material parameter. One can consider the carrier density (n or p) or the carrier mobility (μ_n or μ_p) to be fundamental or microscopic parameters. For a semiconductor material, the resistivity is related to these two parameters (density and mobility) by (20.2). The strength of the Hall effect is to directly determine the sheet carrier density by measuring the voltage generated transversely to the current flow direction in a semiconductor sample when a magnetic field is applied perpendicularly, as shown in Fig. 20.15a. Together with a resistivity measurement technique such as the four-point probe or the van der Pauw technique, Hall measurements can be used to determine the mobility of a semiconductor sample.

In modern semiconductor components and circuits, knowledge of these two fundamental parameters n/p and μ_n/μ_p is critical. Currently, Hall effect measurements are one of the most commonly used characterization tools in the semiconductor industry and research laboratories. This is not just because of the parameters that can be extracted for use in device modeling or materials characterization, but also because of the quantum Hall effect (QHE) in condensed matter physics [20.18]. Moreover, in the applied electronics domain, one should note the development of different sensors based on the physical principle of the Hall effect, such as commercial CMOS Hall sensors.

As is very often the case, the development of a characterization technique is related to its cost, simplicity

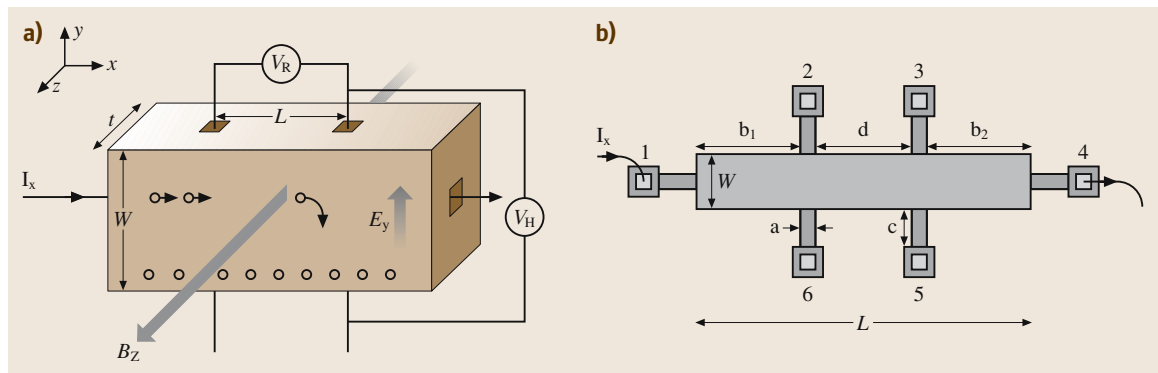


Fig. 20.15 (a) Representation of the Hall effect in an p-type bar-shaped semiconductor. (b) Practical sample geometry: a six-terminal Hall-bar geometry

of implementation and ease of use. Since these practical characteristics are satisfied even when specially shaped samples are required, then the Hall effect measurement technique has become a very popular method of characterizing materials.

In this section, we will first present the physical principle of the Hall effect. Then we will show how it can be used to determine the carrier density and mobility. Finally, the influence of the Hall scattering factor will be presented, followed by some practical issues about the implementation of the Hall effect method.

20.2.1 Physical Principles

The Hall effect was discovered by *Hall* in 1879 [20.19] during an experiment on current transport in a thin metal strip. A small voltage was generated transversely when a magnetic field was applied perpendicularly to the conductor.

The basic principle of this Hall phenomenon is the deviation of some carriers from the current line due to the Lorentz force induced by the presence of a transverse magnetic field. As a consequence, a voltage drop V_H is induced transversely to the current flow. This is shown in Fig. 20.15a for a p-type bar-shaped semiconductor, where a constant current flow I_x in the x -direction and a magnetic field in the z -direction results in a Lorentz force on the holes. If both holes and electrons are present, they deviate towards the same direction. Thus, the directions of electrical and magnetic fields must be accurately specified.

The Lorentz force is given by the vector relation

$$F_L = q(\mathbf{v} \times \mathbf{B}) = -qv_x B_z, \quad (20.33)$$

where v_x is the carrier velocity in the x -direction. Assuming a homogeneous p-type semiconductor

$$v_x = \frac{I}{qtWp}. \quad (20.34)$$

As a consequence, an excess surface electrical charge appears on one side of the sample, and this gives rise to an electric field in the y -direction E_y . When the magnetic force F_L is balanced by the electric force F_{EL} , then the Hall voltage V_H is established, and from a balance between F_L and F_{EL} , we get

$$F = F_L + F_{EL} = -qv_x B_z + qE_y = 0, \quad (20.35)$$

$$\text{so } E_y = \frac{BI}{qtWp}. \quad (20.36)$$

Also, the Hall voltage V_H is given by

$$V_H = V_y = E_y W = \frac{BI}{qt p}. \quad (20.37)$$

So if the magnetic field B and the current I are known, then the measurement of the Hall voltage gives the hole sheet concentration p_s from

$$p_s = pt = \frac{BI}{qV_H}. \quad (20.38)$$

If the conducting layer thickness t is known, then the bulk hole concentration can be determined [see (20.40)] and expressed as a function of the Hall coefficient R_H , defined as

$$R_H = \frac{tV_H}{BI} \quad (20.39)$$

$$\text{and } p = \frac{1}{qR_H} \quad (20.40)$$

Using the same approach for an n-type homogeneous semiconductor material leads to

$$R_H = -\frac{tV_H}{BI}, \quad (20.41)$$

$$\text{and } n = -\frac{1}{qR_H} \quad (20.42)$$

Now, if the bulk resistivity ρ is known or can be measured at the same time using a known sample such as a Hall bar or van der Pauw structure geometry in zero magnetic field, then the carrier drift mobility can be obtained from

$$\mu = \frac{|R_H|}{\rho} \quad (20.43)$$

There are two main sample geometries commonly used in Hall effect measurements in order to determine either the carrier sheet density or the carrier concentration if the sample thickness is known, and the mobility. The first one is the van der Pauw structure presented in Sect. 20.1.1. The second one is the Hall bar structure shown in Fig. 20.15b, where the Hall voltage is measured between contacts 2 and 5, and the resistivity is measured using the four-point probes technique presented in Sect. 20.1.1 (contacts 1, 2, 3 and 4). Additional information about the shapes and sizes of Hall structures can be found in [20.3, 4, 20].

Whatever the geometry used for Hall measurements, one of the most important issues is related to the offset voltage induced by the nonsymmetric positions of the contact. This problem, and also those due to spurious voltages, can be controlled by two sets of

measurements, one for a magnetic field in on direction and another for a magnetic field in the opposite direction.

The Hall effect has also been investigated on specific structures, and an interesting example can be found in reference [20.21], where a Hall bar structure was combined with a double-gate n-SOI MOSFET. This was done in order to understand the mobility behavior in ultra-thin devices and to validate the classical drift mobility extraction method based on current-voltage measurements.

In the Hall effect experiment, the measurement of the Hall coefficient R_H leads to the direct determination of the carrier concentration and mobility. Moreover, the sign of R_H can be used to determine the type of conductivity of the semiconductor sample. If various types of carriers are present, then the expression for R_H becomes more complex and approximations in the limit of low and high magnetic field are necessary (see Chap. 8 of [20.3]).

We have so far discussed the Hall effect on a uniformly doped substrate or single semiconductor layer deposited on an insulating or semi-insulating substrate. In the case of a semiconductor layer deposited on a semi-conducting substrate of opposite doping type, Hall effect measurements can be performed if the space charge region can act as an insulator. In the case of multilayers, the problem is more difficult, but an approximation for transport experiments has been developed for two-layer

structures [20.23] and applied to different MESFET structures, for instance [20.24].

20.2.2 Hall Scattering Factor

The relations presented above are based on an energy-independent scattering mechanism. With this assumption made, the Hall carrier concentration and the Hall mobility are equal to the carrier concentration and the carrier drift mobility. When this assumption is no longer valid, these electrical parameters are different and the Hall scattering factor r_H must be taken into account. In this case (20.40), (20.42) and (20.43) must be modified as follows:

$$p_H = \frac{r_H}{qR_H} = r_H p, \quad (20.44)$$

$$n_H = -\frac{r_H}{qR_H} = r_H n, \quad (20.45)$$

and

$$\mu_H = r_H \frac{|R_H|}{\rho} = r_H \mu. \quad (20.46)$$

The Hall scattering factor [20.25] is related to the energy dependence of the mean free time between carrier collisions $\tau(E)$, and r_H is given by:

$$r_H = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2}. \quad (20.47)$$

According to theory [20.3], the Hall scattering factor tends to unity in the limit of high magnetic field. Therefore, r_H at low magnetic fields can be determined by measuring the Hall coefficient in the limit of both high and low magnetic fields [20.25] using:

$$r_H = \frac{R_H(B)}{R_H(\infty)} \quad (20.48)$$

Depending on the scattering mechanism involved (lattice, ionized or neutral impurity, electron, or phonon scattering), r_H is found to vary between 0.6 and 2 [20.26]. However, due to valence band distortion effects, values as low as 0.26 have been found in strained p-type SiGe epilayers [20.27]. Therefore, the Hall carrier concentration and especially the Hall mobility must be distinguished from carrier concentration and carrier drift mobility.

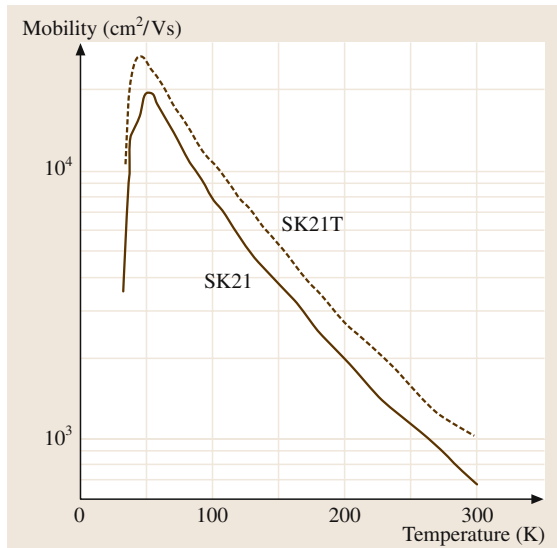


Fig. 20.16 Hall mobility as a function of temperature on two SOI films. After [20.22]

As the different scattering mechanisms have different temperature (T) dependences, then the Hall mobility as function of temperature is often used to separate the different scattering processes. An example is given in Fig. 20.16 for silicon-on-insulator (SOI) films [20.22]. The increase in the mobility between 4–45 K, which is given by $\mu \propto T^{2.95}$, is related to the ionized donor scattering mechanism.

The decrease in mobility between 46–120 K given by $\mu \propto T^{-1.55}$ is associated with lattice scattering. However, after 150 K, the rapid decrease in mobility observed, where $\mu \propto T^{-2.37}$, suggests that other scattering mechanisms as well as the lattice scattering mechanism, such as electron or phonon scattering, must be taken into account.

20.3 Capacitance–Voltage Measurements

Capacitance–voltage (C – V) measurements are normally made on metal–oxide semiconductor (MOS) or metal–semiconductor (MS) structures in order to determine important physical and defect information about the insulator and semiconductor materials. For example, high-frequency (HF) and low-frequency (LF) or quasi-static C – V measurements in these structures are used to determine process and material parameters – insulator thickness, doping concentration and profile, density of interface states, oxide charge density, and work function or barrier height. In this section, we describe various C – V measurements and how they can be used to provide process parameters as well as valuable information about the quality of the materials. A typical C – V curve for a MOS capacitor with an n-type semiconductor is shown in Fig. 20.17. For a MOS capacitor with a p-type substrate, the C – V curve be similar to that in Fig. 20.17, but reflected about the y-axis.

20.3.1 Average Doping Density by Maximum–Minimum High-Frequency Capacitance Method

The maximum–minimum high-frequency (HF) capacitance method uses the HF capacitance under strong accumulation (C_{OX}) and strong inversion ($C_{HF,min}$) to determine the average doping density (see [20.29][pp. 406–408]). Note that under strong inversion and at high frequencies, the interface trap capacitance is negligible ($C_{it} \approx 0$). Under strong inversion, the depletion width (w_{max}) is a maximum and so the high frequency capacitance per unit area $C_{HF,min}$ is a minimum, since the minority carriers cannot respond to the high-frequency signal. Since the inversion layer is very thin compared to the depletion layer, then

$$w_{max} = \epsilon_{Si} \left(\frac{1}{C_{HF,min}} - \frac{1}{C_{OX}} \right), \quad (20.49)$$

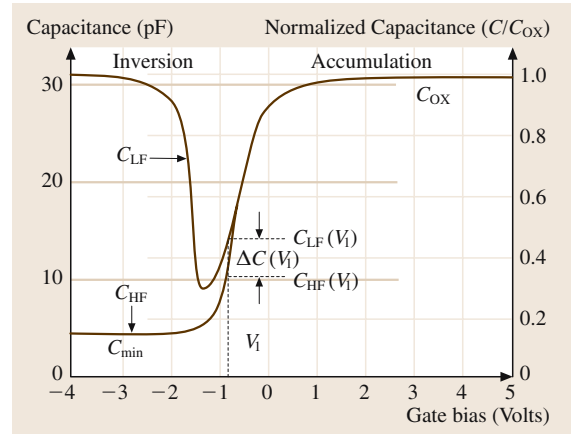


Fig. 20.17 Typical C – V curve for a MOS capacitor on an n-type substrate. After [20.28]

where ϵ_{Si} is the permittivity of silicon and C_{OX} is the gate oxide capacitance per unit area.

At the conditions for w_{max} , the band bending ψ_{max} is a maximum, and it is

$$\begin{aligned} \psi_{max} &= 2\phi_B + \frac{kT}{q} \ln \left(2 \frac{q}{kT} \phi_B - 1 \right) \\ &= 2 \frac{kT}{q} \left\{ \ln \left(\frac{n}{n_i} \right) + \frac{1}{2} \ln \left[2 \ln \left(\frac{n}{n_i} \right) - 1 \right] \right\}, \end{aligned} \quad (20.50)$$

where $\phi_B = (k_B T/q) \times \ln(n/n_i)$ is the shift of the Fermi level from the intrinsic Fermi level $\phi_i = (E_c - E_v)/2q$ in the bulk of the silicon in the MOS structure due to the doping concentration n , and n_i is the thermally generated carrier concentration in silicon. For a uniformly doped sample,

$$w_{max}^2 = \frac{2\epsilon_{Si}\psi_{max}}{qn} \quad (20.51)$$

and from (20.49) and (20.51), a relation between the doping concentration n and the measured capacitance

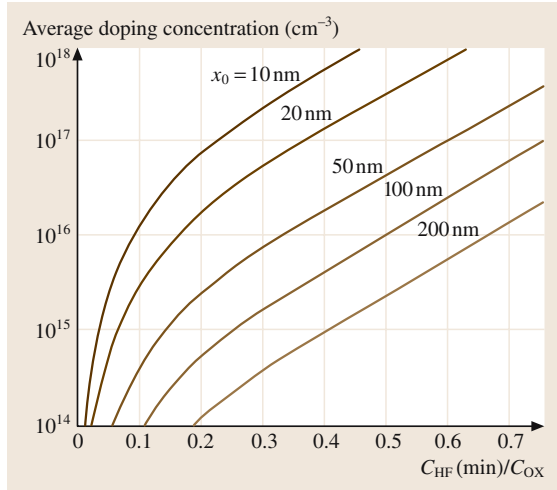


Fig. 20.18 Doping concentration n as function of $C_{\text{HF,min}}/C_{\text{OX}}$ with oxide thickness, based on (20.52). After [20.29]

can be established ([20.29], p. 407) as

$$\frac{n}{\ln\left(\frac{n}{n_i}\right) + \frac{1}{2} \ln\left[2 \ln\left(\frac{n}{n_i}\right) - 1\right]} = \frac{4kT}{q^2 \epsilon_{\text{Si}}} \frac{C_{\text{OX}}^2}{\left(\frac{C_{\text{OX}}}{C_{\text{HF,min}}} - 1\right)^2}. \quad (20.52)$$

Equation (20.52) is a transcendental equation in average doping concentration n that can be solved numerically by iteration. Figure 20.18 shows the solutions as function of $C_{\text{HF,min}}/C_{\text{OX}}$ with oxide thickness, and this can be used to obtain the average doping n graphically. Equation (20.52) can be further simplified by neglecting the term $0.5 \ln[2 \ln(n/n_i) - 1]$, and assuming $C_{\text{OX}} = C_{\text{HF,max}}$ [20.30]. Also, an approximation of (20.52) for the average doping concentration n in unit cm^{-3} is obtained in [20.4] and [20.31] for silicon MOS structures at room temperature, and this is given by

$$\begin{aligned} \log_{10}(n) &= 30.38759 + 1.68278 \\ &\quad \times \log_{10}(C_{\text{DM}} - 0.03177) \\ &\quad \times [\log_{10}(C_{\text{DM}})]^2, \end{aligned} \quad (20.53)$$

where the depletion capacitance (per cm^2 of area) C_{DM} is defined as

$$C_{\text{DM}} = \frac{C_{\text{HF,min}} C_{\text{OX}}}{C_{\text{OX}} - C_{\text{HF,min}}}, \quad (20.54)$$

where all capacitances are in units of F/cm^2 .

20.3.2 Doping Profile by High-Frequency and High-Low Frequency Capacitance Methods

The doping profile in the depletion layer can be obtained ([20.29], Sect. 9.4) by assuming that the depletion capacitance per unit area C_{D} and the oxide capacitance per unit area C_{OX} are connected in series; that is, that the measured high frequency capacitance C_{HF} is given by

$$\frac{1}{C_{\text{HF}}} = \frac{1}{C_{\text{D}}} + \frac{1}{C_{\text{OX}}} \Rightarrow \frac{1}{C_{\text{D}}} = \frac{1}{C_{\text{HF}}} - \frac{1}{C_{\text{OX}}}. \quad (20.55)$$

For a particular gate biasing V_{G} of the MOS structure, the depletion thickness $w(V_{\text{G}})$ is obtained from C_{D} as

$$w(V_{\text{G}}) = \epsilon_{\text{Si}} \left(\frac{1}{C_{\text{HF}}} - \frac{1}{C_{\text{OX}}} \right). \quad (20.56)$$

The doping concentration $n(V_{\text{G}})$ is given by the slope of the $(1/C_{\text{HF}})^2$ versus V_{G} characteristic, given by

$$n(w) = \frac{-2}{q \epsilon_{\text{Si}} \frac{\partial}{\partial V_{\text{G}}} \left(\frac{1}{C_{\text{HF}}^2} \right)}. \quad (20.57)$$

Note that a plot of $1/C_{\text{HF}}^2$ versus V_{G} (Fig. 20.19) can yield important information about the doping profile. The average n is related to the reciprocal of the slope in the linear part of the $1/C_{\text{HF}}^2$ versus V_{G} curve, and the intercept with V_{G} at a value of $1/C_{\text{OX}}^2$ is equal to

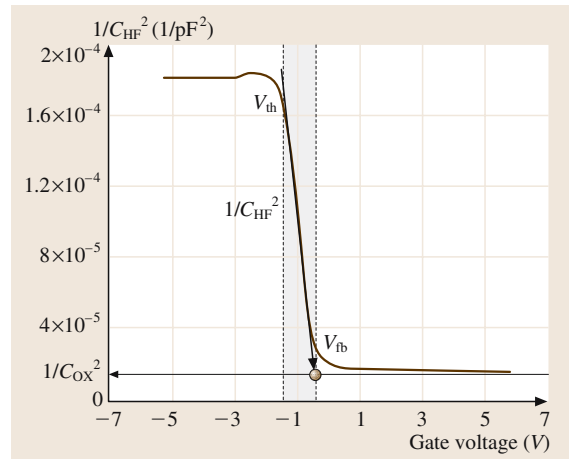


Fig. 20.19 A $1/C_{\text{HF}}^2$ versus V_{G} plot [20.30]. The slope of the fitted arrow line is proportional to the average doping, and the arrow points to the flat-band voltage V_{FB} , obtained at the V intercept with $1/C_{\text{OX}}^2$, shown with the second horizontal arrow

the flat-band voltage V_{FB} caused by the fixed surface charge Q_{SS} and the gate–semiconductor work function ψ_{MS} [20.3, 30].

Equation (20.57) does not take into account the impact of interface traps, which cause the C – V curve to stretch. The traps are slow and do not respond to the high frequency of the test signal, but they do follow the changes in the gate bias. Therefore, ∂V_G must be replaced with ∂V_{G0} in (20.57), with ∂V_{G0} representing the case when no interface traps are present.

The value of ∂V_{G0} can be obtained by comparing high- and low-frequency (quasi-static) C – V curves for a MOS structure at the same gate biases V_G . Therefore, the ratio $\partial V_{G0}/\partial V_G$ can be found at any gate bias V_G , since the band-bending is the same for both HF and LF capacitances. In [20.29] (Sect. 9.4), it is shown that

$$\frac{\partial V_{G0}}{\partial V_G} = \frac{C_{OX} + C_D}{C_{OX} + C_D + C_{IT}} = \frac{1 - C_{LF}/C_{OX}}{1 - C_{HF}/C_{OX}}. \quad (20.58)$$

In this case (20.57) is modified to

$$n(w) = \frac{-2}{q\epsilon_{Si} \frac{\partial}{\partial V_G} \left(\frac{1}{C_{HF}} \right)} \frac{1 - C_{LF}/C_{OX}}{1 - C_{HF}/C_{OX}} \quad (20.59)$$

as originally proposed in [20.32] and illustrated in Fig. 20.20. As seen from Fig. 20.20, the stretching of the C – V curves due to the interface states induced by stress in Fig. 20.20a causes a disparity in the doping profile in Fig. 20.20b if only the high frequency capacitance is used. The disparity is well-suppressed in Fig. 20.20c by the high–low frequency capacitance measurement, taking into account the stretching of the C – V characteristics. Provided that the depletion layer capacitance is

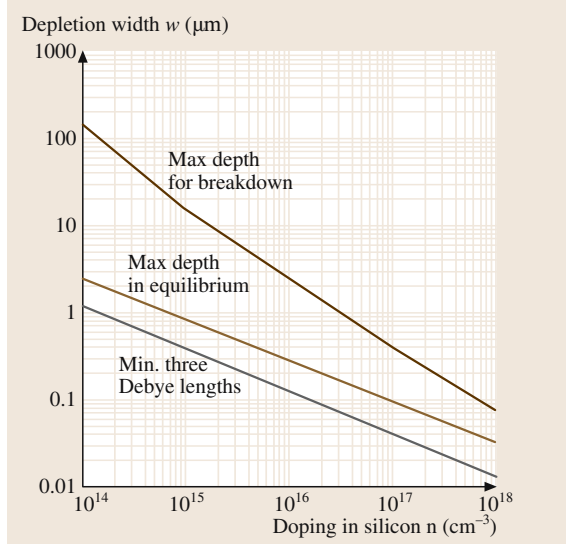


Fig. 20.21 Limitations on the depth achievable when profiling the doping of silicon MOS structures via C – V measurements at room temperature (After [20.4] p. 86)

measured at a high frequency, the depletion layer width w is still obtained by (20.56).

Note that the maximum depth w_{max} (20.51) and the resolution Δ_w of the doping profile by means of C – V measurements is limited by the maximum band-bending ψ_{max} and the extrinsic Debye length λ_{Debye} , given by (20.50) and (20.60), respectively, and λ_{Debye} is

$$\lambda_{Debye} = \sqrt{\frac{\epsilon_{Si} k T}{q^2 n}}. \quad (20.60)$$

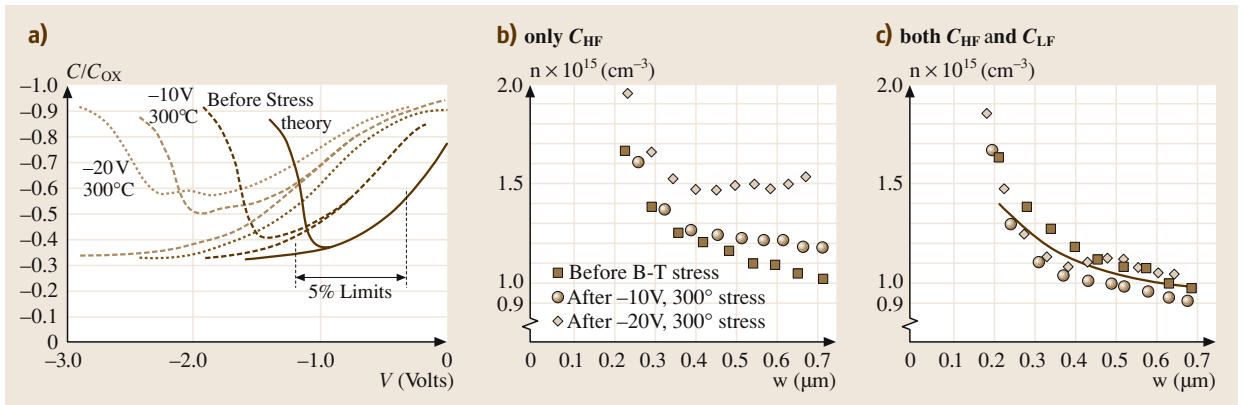


Fig. 20.20a–c C – V curves and doping profiles of a MOS structure with 145 nm oxide and uniform doping of 10^{15} cm^{-3} before and after bias temperature stress [20.32]. (a) Theoretical and (stretched) measured C – V curves. (b) Doping profile deduced from only C_{HF} (see (20.57)); (c) Doping profile deduced from both C_{HF} and C_{LF} (see (20.59)). After [20.32]

The doping profile obtained in this way is reliable for depths w of between $3\lambda_{\text{Debye}}$ and $w_{\text{max}}/2$, when the MOS structure is in depletion and weak inversion, but not in accumulation. That is, $C_{\text{LF}} < 0.7C_{\text{OX}}$ as a simple rule. As illustrated in Fig. 20.21, the range of w values between $3\lambda_{\text{Debye}}$ and equilibrium, obtained via quasi-static C - V measurements, cover about half-a-decade. With proper corrections, the lower distance decreases to one Debye length [20.30]. Using nonequilibrium (transient) C - V measurements in deep depletion, the profiling can be extended to higher distances by about an order of magnitude, but further limitations can appear due to the high-frequency response of the interface charge, measurement errors, avalanche breakdown in deep depletion, or charge tunneling in highly doped substrates and thin oxides. More details are presented in [20.29].

20.3.3 Density of Interface States

Interface traps change their charge state depending on whether they are filled or empty. Because interface trap occupancy varies with the slow gate bias, stretching of the C - V curves occurs, as illustrated in Fig. 20.20. A quantitative treatment of this “stretch-out” can be obtained from Gauss’ law as

$$C_{\text{OX}}(V_G - \psi_S) = -Q_S - Q_{\text{IT}} = -Q_T, \quad (20.61)$$

where Q_S and Q_{IT} are the surface and interface trap charges (per unit area), which are both dependent on the surface band-bending ψ_S , $Q_T = Q_S + Q_{\text{IT}}$ is the total charge in the MOS structure, C_{OX} is the gate capacitance (per unit area), and V_G is the bias applied at the gate of the MOS structure. For simplicity, the (gate metal)-to-(semiconductor bulk) potential ψ_{MS} is omitted in (20.61), but in a real structure the constant ψ_{MS} must be subtracted from V_G . As follows from (20.61), small changes ∂V_G in gate bias cause changes $\partial \psi_S$ in the surface potential bending, and the surface C_S and interface trap C_{IT} capacitances (both per unit area) can represent Q_S and Q_{IT} , given by

$$C_{\text{OX}}\partial V_G = (C_{\text{OX}} + C_S + C_{\text{IT}}) \cdot \partial \psi_S. \quad (20.62)$$

C_S and C_{IT} are in parallel and in series with the C_{OX} , respectively. Therefore, the measured low-frequency capacitance C_{LF} (per unit area) of the MOS structure becomes

$$\begin{aligned} C_{\text{LF}} &= \frac{\partial Q_T}{\partial V_G} = \frac{\partial Q_T}{\partial \psi_S} \cdot \frac{\partial \psi_S}{\partial V_G} \\ &= \frac{C_{\text{OX}}(C_S + C_{\text{IT}})}{C_{\text{OX}} + C_S + C_{\text{IT}}}. \end{aligned} \quad (20.63)$$

Equation (20.63) shows that stretch-out in the C - V curve can arise due to a non-zero value of C_{IT} , which deviates from the ideal case of $C_{\text{IT}} = 0$.

According to [20.29] (p. 142), D_{IT} is the density of interface states per unit area (cm^2) and per unit energy (1 eV) in units of $\text{cm}^{-2}\text{eV}^{-1}$. Since the occupancy of the interface states has a Fermi–Dirac distribution, then upon integrating over the silicon band-gap, the relation between C_{IT} and D_{IT} is

$$C_{\text{IT}}(\psi_S) = qD_{\text{IT}}(\phi_B + \psi_S), \quad (20.64)$$

where $\phi_B = (k_B T/q) \ln(n/n_i)$ is the shift of the Fermi level from the intrinsic level $\phi_i = (E_c - E_v)/2q$ in the silicon bulk of the MOS structure due to the doping concentration n , and n_i is the thermally generated carrier concentration in silicon. Since the derivative of the Fermi–Dirac distribution is a sharply peaking function, then $C_{\text{IT}}(\psi_S)$ at particular ψ_S probes $D_{\text{IT}}(\phi_B + \psi_S)$ over a narrow energy range of $k_B T/q$, in which D_{IT} can be assumed to be constant and zero outside this interval. Thus, varying the gate bias V_G , and therefore ψ_S , (20.64) can be used to obtain the density of states D_{IT} at a particular energy shift $q(\phi_B + \psi_S)$ from the silicon intrinsic (mid-gap) energy E_i .

It is evident from (20.63) and (20.64) that the experimental values for D_{IT} can be obtained only when C_{IT} , and ψ_S are determined from C - V measurements. The simplest way to determine ϕ_B is to get the average doping density n using the maximum–minimum high-frequency capacitance method (see (20.52) and Fig. 20.18), or to use the values of n from doping profiles at $0.9w_{\text{max}}$ - see (20.58) [20.30]. Either the high-frequency or the low-frequency C - V measurement can be used to obtain C_{IT} , but it is necessary to calculate C_S as function of ψ_S , which makes it difficult to process the experimental data.

The most suitable technique for experimentally determining D_{IT} is the combined high–low frequency capacitance method ([20.29], Sect. 8.2.4, p. 332). The interface traps respond to the measurement of low–frequency capacitance C_{LF} , whereas they do not respond to the measurement of the high-frequency measurement C_{HF} . Therefore, C_{IT} can be obtained from measurements by “subtracting” C_{HF} from C_{LF} , given by

$$\begin{aligned} C_{\text{IT}} &= \left(\frac{1}{C_{\text{LF}}} - \frac{1}{C_{\text{OX}}} \right)^{-1} \\ &\quad - \left(\frac{1}{C_{\text{HF}}} + \frac{1}{C_{\text{OX}}} \right)^{-1}. \end{aligned} \quad (20.65)$$

Denoting $\Delta C = C_{\text{LF}} - C_{\text{HF}}$, the substitution of (20.65) into (20.64) provides a direct estimate of D_{IT} from C - V

measurements (see also [20.3], p. 371) as

$$D_{IT} = \frac{\Delta C}{q} \left(1 - \frac{C_{LF}}{C_{OX}}\right)^{-1} \left(1 - \frac{C_{HF}}{C_{OX}}\right)^{-1} \quad (20.66)$$

Note that the combined high–low frequency capacitance method provides C_{IT} and D_{IT} as function of gate bias V_G . However, if D_{IT} needs to be plotted as a function of the position in the energy band-gap, the surface band-bending ψ_S must also be determined as function of gate bias V_G , as follows from (20.64).

There are several ways to obtain the relation between ψ_S and V_G . One way is to create a theoretical plot of C_{HF} versus ψ_S and then, for any choice of C_{HF} , a pair of values for ψ_S and V_G is found (see [20.29] p. 327). This method is relatively simple if the doping concentration n in the silicon is uniform and known, because the high-frequency silicon surface capacitance C_S under depletion and accumulation is a simple function of the band-bending ψ_S , and the flat-band capacitance C_{FB} ([20.29] pp. 84, 97, 164) is given by

$$C_{FB} = \frac{\varepsilon_{Si}}{\lambda_{Debye}} = \sqrt{\frac{\varepsilon_{Si} q^2 n}{k_B T}} \quad (20.67)$$

$$C_S(\psi_S) \approx \begin{cases} \frac{C_{FB}}{\sqrt{2}} \frac{\exp\left(\frac{q\psi_S}{k_B T}\right) - 1}{\sqrt{\exp\left(\frac{q\psi_S}{k_B T}\right) - \frac{q\psi_S}{k_B T} - 1}}, & \psi_S > 0 \text{ in accumulation} \\ C_{FB}, \psi_S = 0 \text{ at flat band,} \\ \frac{C_{FB}}{\sqrt{2}} \frac{1 - \exp\left(\frac{q\psi_S}{k_B T}\right)}{\sqrt{\exp\left(\frac{q\psi_S}{k_B T}\right) - \frac{q\psi_S}{k_B T} - 1}}, & \psi_S < 0 \text{ in depletion.} \end{cases} \quad (20.68)$$

Since C_S is in series with C_{OX} , then the theoretical C_{HF} is obtained as a function of the band-bending ψ_S by

$$\frac{1}{C_{HF}(\psi_S)} = \frac{1}{C_{OX}} + \frac{1}{C_S(\psi_S)} \quad (20.69)$$

For a uniformly doped silicon with SiO_2 as the insulator, the ratio $C_{HF}(V_{FB})/C_{OX}$ at gate bias for flat-band conditions is given ([20.3], p. 349) by

$$\frac{C_{HF}(V_{FB} \text{ or } \psi_S = 0)}{C_{OX}} = \frac{1}{1 + \frac{136\sqrt{T/300}}{t_{ox}\sqrt{n}}}, \quad (20.70)$$

where t_{ox} is the oxide thickness (cm), n is the doping (cm^{-3}), and the T is the temperature (K).

It was demonstrated in [20.29] that the method of using a theoretical plot to obtain the relation between ψ_S and V_G works well in the case of uniformly doped silicon even if only high-frequency C – V measurements are used to obtain the density of states; that is, the $1/C_{HF}^2$ versus V_G plot is almost a straight line. However, with substrates that are not uniformly doped, the method is inconvenient because the corrections in (20.67) and (20.68) are difficult to implement. Therefore, in practice, a method based on low-frequency C – V measurement is preferred [20.30].

Low-frequency C – V measurement was first used to obtain the relation between ψ_S and V_G [20.34]. This method is based on the integration of (20.65) from an initial gate bias V_{G0} , arbitrarily chosen either under strong accumulation or strong inversion, to the desired V_G at which the band-bending $\psi_S(V_G)$ is to be obtained. Since C_{IT} is part of (20.65), then the low-frequency C – V curve ([20.29] p. 93) is integrated as

$$\psi_S(V_G) = \psi_{S0} + \int_{V_{G0}}^{V_G} \left(1 + \frac{C_{LF}(V_G)}{C_{OX}}\right) dV_G. \quad (20.71)$$

The value of ψ_{S0} is selected such that $\psi_S(V_{FB}) = 0$ when integrating from V_{G0} to the flat-band gate voltage V_{FB} . In this case, V_{FB} is usually obtained beforehand from the point of V -intercept with $1/C_{OX}^2$ when extrapolating

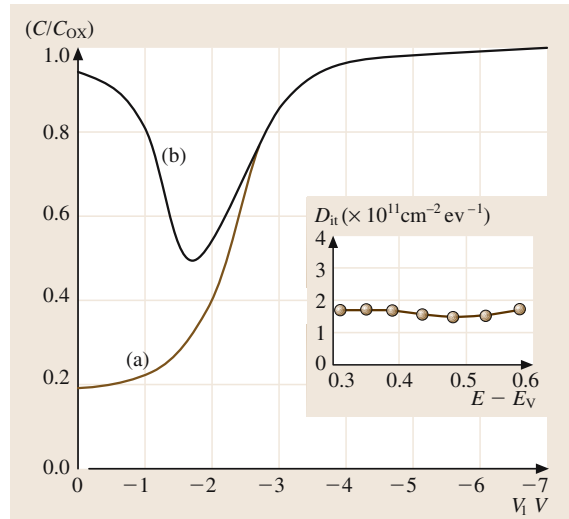


Fig. 20.22a,b Results from the combined high–low frequency capacitance method [20.33]. (a) High-frequency C – V curve; (b) low-frequency C – V curve. The energy profile for the density of interface states D_{IT} is shown in the inset, as calculated by (20.64), (20.66) and (20.71)

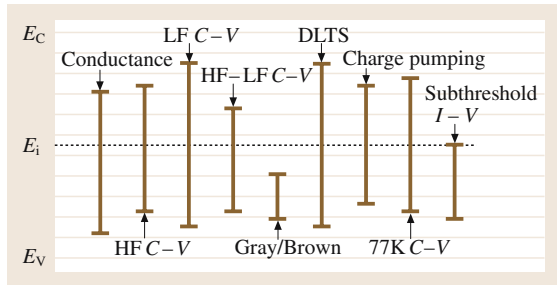


Fig. 20.23 Energy ranges in the silicon band-gap of a p-type substrate over which the density of interface traps can be determined using various measurement methods and characterization techniques. (After [20.3], p. 104)

the linear part of the $1/C_{\text{HF}}^2$ versus V_G curve toward the V_G axis (Fig. 20.19). After determining ψ_{S0} , (20.71) provides the relation between ψ_S and V_G . Thus, the density of states D_{IT} obtained from (20.66) as function of V_G using the combined high–low frequency capacitance method can be plotted against the position of D_{IT} in the silicon band gap, as given by (20.64). High and low frequency $C-V$ measurements can therefore be used to plot the data, as illustrated in the insert of Fig. 20.22.

20.4 Current–Voltage Measurements

20.4.1 $I-V$ Measurements on a Simple Diode

Current–voltage measurements of mainstream semiconductor devices are perhaps the simplest and most routine measurements performed, and they can provide valuable information about the quality of materials used. For example, if we consider the $I-V$ characteristics of a p–n diode structure, the source–substrate or drain–substrate junctions can provide useful information on the quality of the junction, such as whether defects are present (they give rise to generation–recombination currents or large parasitic resistances for the contacts at the source, drain or substrate terminals). This is easily seen from the current–voltage relation given by the sum of the diffusion (I_{DIFF}) and recombination (I_{GR}) currents:

$$I_D = I_{\text{DIFF}} + I_{\text{GR}} = I_{D0} \left[\exp \left(\frac{eV_D}{nk_B T} \right) - 1 \right] + I_{\text{GR0}} \cdot \exp \left(\frac{eV_D}{2k_B T} \right) \quad (20.72)$$

Overall, many different techniques are used to determine the density of states D_{IT} (please see Fig. 20.13). For some of these techniques, the ability to sense the energy position of D_{IT} in the band-gap of silicon is summarized in [20.3]. Most of them use $C-V$ measurements, but others are based on $I-V$ measurements taken during the subthreshold operation of MOS transistors, deep-level transient spectroscopy (DLTS), charge pumping (CP) in a three-terminal MOS structure, cryogenic temperature measurements, and so on. Each technique has its strengths and weaknesses, which are discussed in [20.3, 35].

In the methods discussed above, it has been assumed that the gate bias V_G varies slowly with time, 20–50 mV/s, and that the MOS structure is in equilibrium; that is, the minority carriers are generated and the inversion layer is readily formed in the MOS structure when V_G is above the threshold. However, the time constant for minority carrier generation is high in silicon (≈ 0.1 s or more), and it is possible to use nonequilibrium high-frequency $C-V$ measurements to further analyze the properties of the MOS structure. Some applications of these methods are presented later.

where I_{D0} and I_{GR0} are the zero-bias diffusion and recombination currents respectively, n is an ideality factor (typically 1), and V_D is the voltage across the intrinsic diode, which is given by

$$V_D = V_{\text{applied}} - I_D R_{\text{parasitic}}. \quad (20.73)$$

For (20.72), a plot of $\ln(I_D)$ versus V_D allows us to separate out the diffusion and the recombination current components. From (20.72) and (20.73), we can also use the diode's $I-V$ characteristics to determine the parasitic resistance in series with the intrinsic diode, as described in detail in [20.36]. In most cases, this $R_{\text{parasitic}}$ is dominated by the contact resistance.

20.4.2 $I-V$ Measurements on a Simple MOSFET

Simple current–voltage measurements – drain current versus gate voltage ($I_{\text{DS}} - V_{\text{GS}}$), and I_{DS} versus drain voltage (V_{DS}) – are routinely taken on MOSFETs in order to study their electrical characteristics; however, these can also be used to obtain useful information

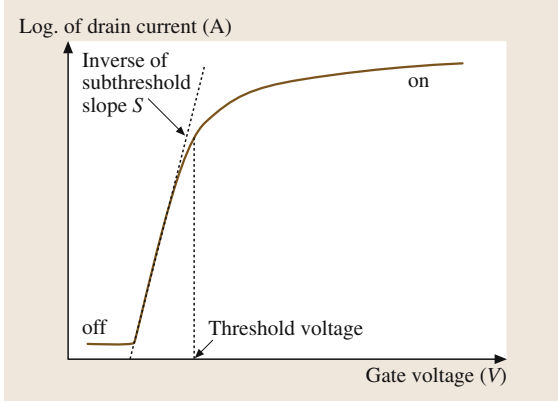


Fig. 20.24 Typical subthreshold characteristics of a MOS-FET. The interface state density can be extracted from S

on the quality of the semiconductor, contacts, oxide and semiconductor–oxide interface. For example, the $I_{DS} - V_{GS}$ characteristics at very small V_{DS} biases (linear region of operation) for a set of test transistors of fixed channel width and different channel lengths is often used to extract parameters such as the threshold voltage (V_T), the transconductance (g_m), the intrinsic mobility (μ_0) and the mobility degradation coefficients θ_0 and η , the parasitic source (R_S) and the drain resistances (R_D) in series with the intrinsic channel, the channel length reduction ΔL , the output conductance (g_{DS}) and the subthreshold slope (S) [20.37].

These parameters are required for modeling and they directly impact the device's performance. However, some of these parameters can also be used to assess the quality of the silicon–silicon dioxide (Si–SiO₂) interface [20.38, 39]. For example, interface states at the Si–SiO₂ interface can change the threshold voltage, the subthreshold slope and the mobility, all of which will impact on the drain–source (I_{DS}) current flowing through the device. Here, we look at one parameter in more detail – the subthreshold slope S in mV (of V_{GS})/decade (of I_{DS}).

First, the interface trap density (D_{IT}) can be determined from a semi-log plot of $I_{DS} - V_{GS}$ characteristics at very low drain biases, as shown in Fig. 20.24. We start with the expression for the subthreshold slope

$$S = \frac{k_B T}{q} \ln(10) \left(1 + \frac{C_D + C_{IT}}{C_{OX}} \right) \quad (20.74)$$

in which

$$C_D = \frac{q \epsilon_{Si} N_A}{\sqrt{2\psi + |V_B| - k_B T/q}} \quad (20.75)$$

D_{IT} can then be calculated from

$$C_{IT} = q D_{IT} \quad (20.76)$$

once C_{IT} is determined from (20.74). In fact, a recent comparison in [20.40] of the interface trap densities extracted from capacitance, subthreshold and charge pumping measurements produced similar results, demonstrating that simple and fast $I-V$ measurements based on the subthreshold technique can provide useful information on the Si–SiO₂ quality.

20.4.3 Floating Gate Measurements

The floating gate technique is another simple $I-V$ measurement in which the evolution of the drain current I_{DS} is monitored after the gate bias has been removed. It has been shown to be particularly useful when monitoring early-mode hot-carrier activity in MOS transistors [20.40, 41]. In this measurement, we first check the oxide quality by biasing the transistor in the strong linear region (very low V_{DS} and a V_{GS} well above V_T), and then lift the gate voltage probe so that $V_{GS} = 0$ V and measure the evolution of I_{DS} with time. For a high-quality gate and spacer oxide, I_{DS} remains constant for a long time, indicating that there is negligible carrier injection across the gate oxide through Fowler–Nordheim tunneling or other leakage mechanisms.

A second precaution is to have a dry or inert gas such as nitrogen flowing over the chip to reduce the possibility of other leakage mechanisms such as that from water vapor. The measurement set-up for this experiment is shown in Fig. 20.25. The

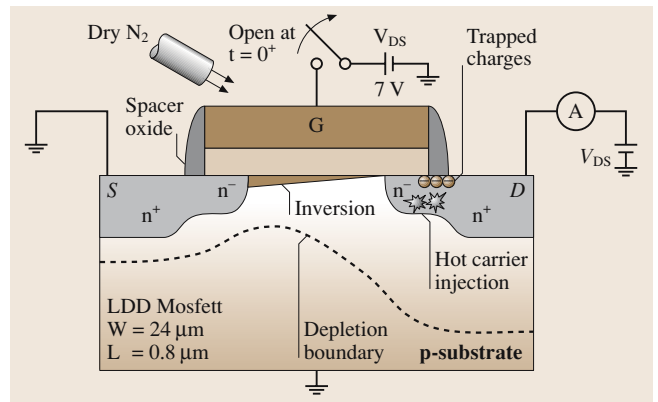


Fig. 20.25 Schematic diagram showing the set-up used for floating gate measurements. The area where charge is trapped after hot electrons are applied is shown

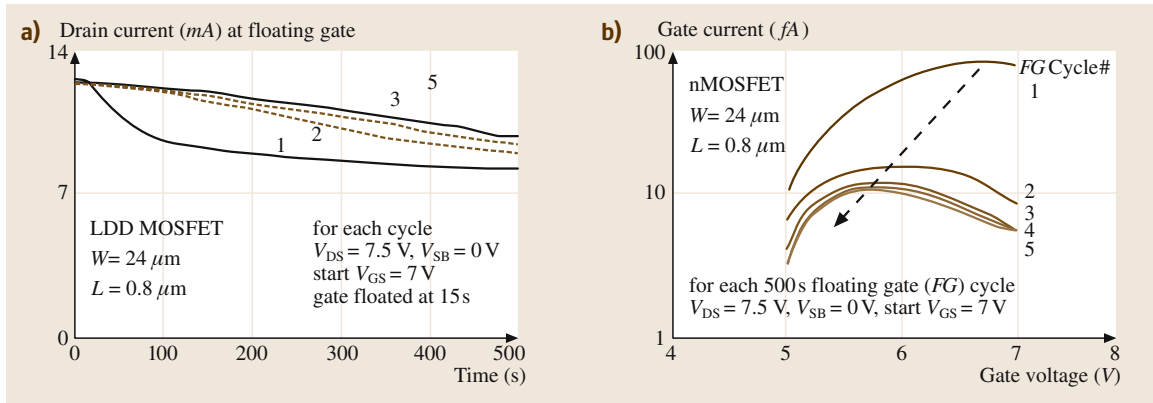


Fig. 20.26 (a) Evolution of the drain current over five floating gate cycles. Biasing conditions were chosen to maximize hot-electron gate currents. Note that the maximum drop in drain current occurs after the first floating gate cycle. (b) Extracted gate currents using (20.77) and the measurements in (a). As with the drain current, the maximum change in gate current occurs between floating gate cycles 1 and 2. The shift in the peak of the gate current is explained in [20.38,41].

evolution of the floating gate current over several cycles with the gate voltage applied and then removed is shown in Fig. 20.26. The biasing voltages and time at which the gate is floated are also given in Fig. 20.26a. For this experiment, a biasing condition of $V_{GS} \approx V_{DS}$ was chosen for a high-impact ionization-induced gate current, but a lower-than-maximum electron injection situation was used for the initial biasing condition.

From the evolution of I_{DS} and the $I_{DS} - V_{GS}$ characteristics of a virgin (not intentionally stressed) transistor at the same V_{DS} as the floating gate measurements, and from measurements of the total capacitance associated with the gate (C_G), the gate current (I_G) evolution after each floating gate cycle can be deter-

mined using

$$I(V_{GS}) = C_G \frac{dV_{GS}}{dt} \quad (20.77)$$

This $I_G - V_{GS}$ evolution is shown in Fig. 20.26b. An ancillary benefit of the floating gate technique is that very small gate currents (in the fA range or even smaller) can be easily determined by measuring much larger drain currents using, for example, a semiconductor parameter analyzer. The reason for this is that the gate current is not directly measured in this technique – it is determined from $I_{DS} - V_{GS}$ and (20.77). Also, the change in the floating gate current after the first few cycles can be used to monitor for early mode failure after statistical evaluation.

20.5 Charge Pumping

Charge pumping (CP) is another electrical technique that is well suited to studying semiconductor–insulator interfaces in MOSFETs [20.42–47]. There are several versions of the CP technique: spatial profiling CP [20.43–47], energy profiling CP [20.48], and, more recently, new CP techniques [20.49] that permit the determination of both interface states (N_{IT}) and oxide traps (N_{OT}) away from the interface and inside the oxide. The charge pumping technique is more complicated than either of the $I-V$ or floating gate methods. However, it is a very powerful technique for assessing interface quality and it works well even with very small transistor geome-

tries and very thin gate oxides, where tunneling can be a problem.

The charge pumping technique was first used in 1969 [20.50] to measure the interface traps at Si–SiO₂ interface. Since then, there have been numerous publications with enhancements, refinements and applications of the technique to a variety of semiconductor–insulator interfacial studies. In the basic charge pumping experiment, the gate of an NMOST (for example) is pulsed from a low value (V_L) when the device is in accumulation to a high value (V_H) when the device is in inversion. This results in the filling of traps between

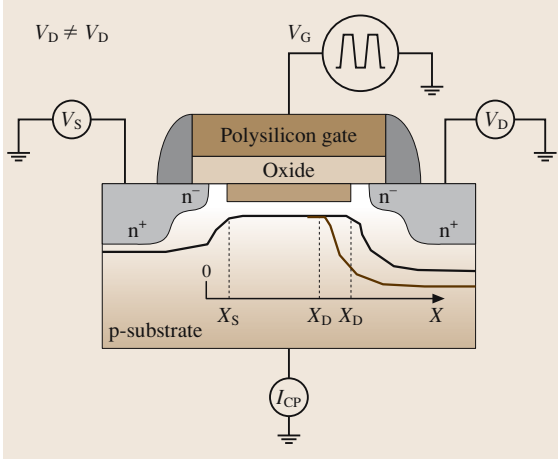


Fig. 20.27 Example of spatial profiling charge-pumping set-up used when the source and drain biases are slightly different

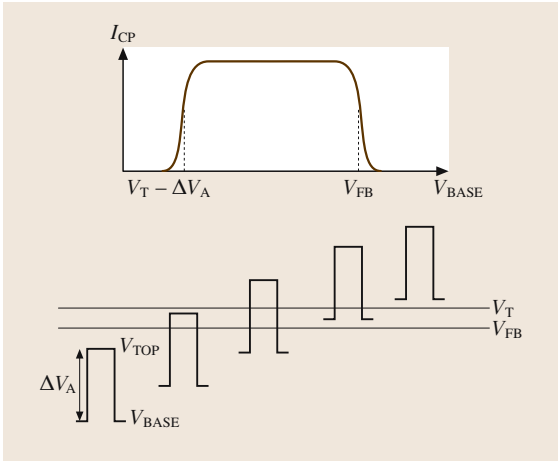


Fig. 20.28 Demonstration of how the CP curve is generated by varying the base level of the pulse so that the entire pulse is between V_{FB} and V_T

$E_{F,ACC}$ (corresponding to V_L) and $E_{F,INV}$ (corresponding to V_H) with holes and electrons, respectively. When pulsing the gate between accumulation at V_L and inversion at V_H , a current flows due to the repetitive recombination at the interface traps of minority carriers from the source and drain junctions with majority carriers from the substrate. This current is termed the charge pumping current, and it was found to be proportional to the frequency of the gate pulse, the gate area and the interface state density. Its sensitivity is better than $10^9 \text{ cm}^{-2} \text{ eV}^{-1}$. In the traditional CP experiment

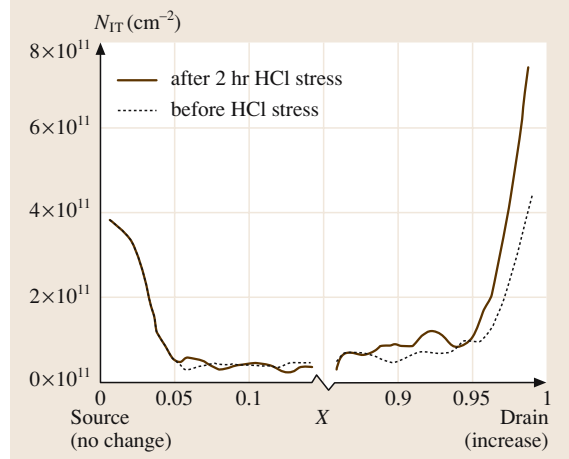


Fig. 20.29 Spatial interface state distribution over the channel in a $1 \mu\text{m}$ -long device. The stress was applied for 2 h at $V_{DS} = 5 \text{ V}$ and $V_{GS} = 2.4 \text{ V}$

shown in Fig. 20.27, but with $\Delta V_S = |V_D - V_S| = 0 \text{ V}$, the gate G is connected to a pulse generator, a reverse bias V_R or no bias is applied to both sources S and drain D terminals, and the charge pumping current flowing in the substrate terminal, I_{CP} , is measured. To generate a typical charge pumping curve (as shown in the top part of Fig. 20.28), the base level of the pulse is varied, taking the transistor from below flat-band to above surface inversion conditions, as shown in the bottom part of Fig. 20.28.

In the traditional CP experiment, the charge pumping current I_{CP} is given by

$$I_{CP} = q A_{GATE} \overline{D_{IT}} \Delta E, \quad (20.78)$$

where

$$\Delta E = (E_{F,INV} - E_{F,ACC}). \quad (20.79)$$

This expression assumes that the electrical and physical channel lengths are the same. However, for short channel devices, this assumption results in an error. Therefore, a more accurate expression is

$$I_{CP} = q f W \int_{x_s}^{x_d} N_{IT}(x) \cdot (q \Delta \psi_{SO}) \quad (20.80)$$

$$x_d = L - \sqrt{\frac{2\epsilon_{Si}}{qN}} (\sqrt{V_D + \psi_S} - \sqrt{\psi_S}). \quad (20.81)$$

The interface state density at the edge of the drain depletion region [$N_{IT}(x_d)$] is given by

$$N_{IT}(x_d) = \frac{1}{qfWq\Delta\psi_{s0}} \left(\frac{dx_s}{dV_S} \right)^{-1} \cdot \left(\frac{dI_{CP}}{dV_S} \right)_{V_D=\text{constant}} \quad (20.82)$$

When performing spatial profiling CP experiments, some precautions are required. The first is that a voltage difference ΔV_S between V_D and V_S that is too small results in a difference in I_{CP} that is too small as well, and hence a large error in $N_{IT}(x)$, as indicated from (20.82).

On the other hand, values of ΔV_S that are too large result in a I_D that is too high and hence more substrate current I_B . This current I_B can interfere with I_{CP} if ΔV_S is large or if L is very short, resulting in a large error in $N_{IT}(x)$. The range 50–100 mV for ΔV_S seems to be a good compromise for the devices investigated in [20.43, 44]. Experimental results for spatial profiling CP measurements indicate that $N_{IT}(x)$ peaks near S/D edges. However, after normal mode stress, $N_{IT}(x)$ only increases near D. This is shown in Fig. 20.29. More details about charge pumping can be found in a recent review [20.51].

20.6 Low-Frequency Noise

20.6.1 Introduction

Low-frequency noise (LFN) spectroscopy requires very good experimental skills in the use of low-noise instrumentation as well as grounding and shielding techniques. Other special considerations are also required, which are discussed later. Although it is time-consuming to perform, it has been widely used to probe micro-

scopic electrical transport in semiconductors and metals. LFN is very sensitive to defects in materials and devices, and large differences in LFN characteristics can be observed in devices with identical electrical current–voltage characteristics. This is mainly because electrical I – V measurements only probe the average or macroscopic transport in devices and so are not as sensitive to defects as LFN. Due to its sensitivity to defects, traps or generation–recombination centers, LFN has been proposed as a good tool for predicting device reliability. For example, LFN has been used to predict the reliabilities of metal films [20.52], and has been used in processing steps that produce photodetectors with better performance [20.53, 54]. LFN noise is sensitive to both bulk and surface defects or contaminants of a material.

Using low-frequency noise spectroscopy and biasing the transistor in saturation, we can spatially profile the defect density near the drain and source terminals for devices in normal and reverse modes of operation [20.55]. Low-frequency noise in the linear region also allows us to extract the average defect density over the entire channel region at the silicon–silicon dioxide interface [20.56, 57]. Noise experiments were performed on small-geometry polysilicon emitter bipolar transistors to investigate the number of interface states in the thin interfacial oxide layer between the monocrystalline and polycrystalline silicon [20.58–68]. Recent experiments using body or substrate bias (V_B) in a MOS transistor allowed us to look at the contribution of bulk defects (defects away from the silicon–silicon dioxide interface) and their contribution to device noise [20.69–71]. This is important since substrate biasing has been proposed as a means to cleverly manage power dissipation and speed in emerging circuits and systems [20.72].

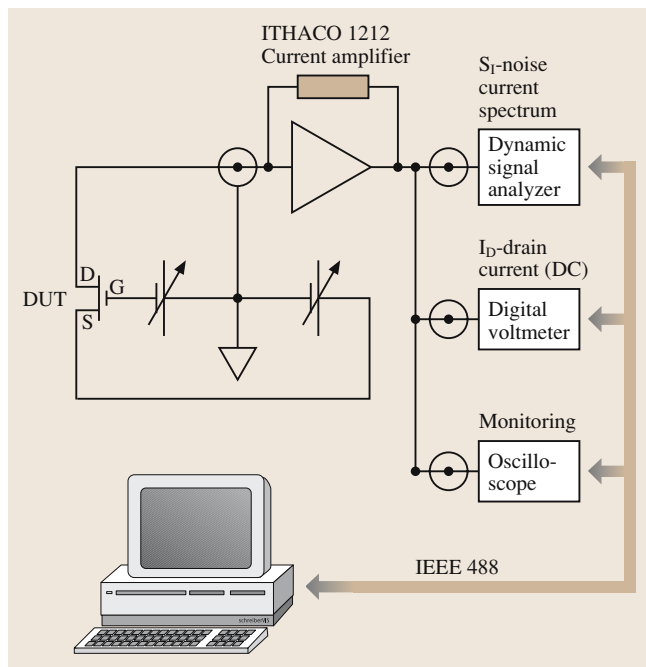


Fig. 20.30 System for measuring low-frequency noise. In this diagram, the device under test is any field-effect or thin-film transistor

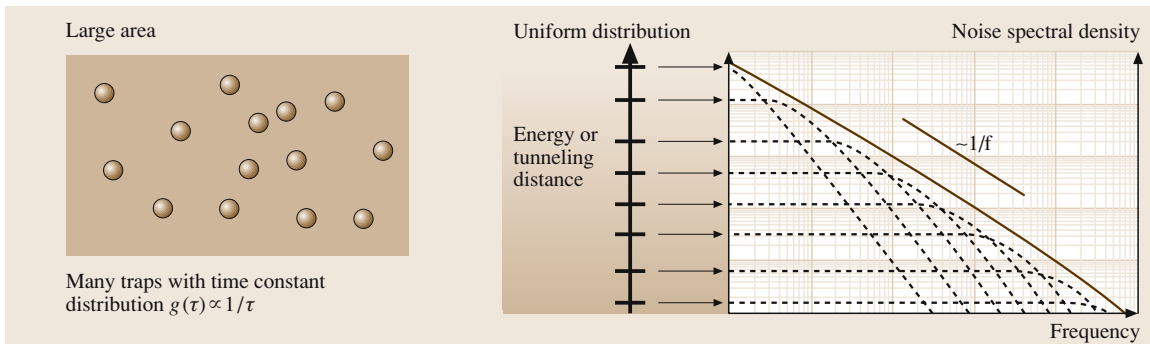


Fig. 20.31 Schematic representation of a “large”-area PE-BJT with many traps distributed uniformly across the band gap and the emitter area, and with a $g(\tau) = 1/\tau$ distribution for the time constant. The resulting spectrum is $1/f$ noise

We will discuss how low-frequency noise (LFN) spectroscopy can be applied to the interfacial oxide layer between the mono-silicon and polysilicon emitter in bipolar junction transistors (BJTs) here. The experimental system shown in Fig. 20.30 is used for LFN measurements of FETs; the same system can also be used for BJTs.

As mentioned before, special attention must be paid to grounding and shielding in LFN measurements, as this is crucial to minimizing the effects of experimental and environmental noise sources on the device under test (DUT). Because electric power supplies are noisy, especially at 60 Hz (in North America) and its harmonics, and this noise can dominate the noise of the DUT, batteries are often used to supply the voltage. Metal film resistors are the preferred means of changing the biasing conditions, because of their better low-noise characteristics compared to carbon resistors, for example.

With these experimental precautions taken, the noise signal from the transistor might still be too low to be directly measured using a spectrum or signal analyzer.

Therefore, a low-noise voltage or current amplifier, whose input noise sources are lower than that of the noise signal, is used to boost the noise signal. In addition, other instruments might be used to measure currents or voltages, or to display the waveforms (as shown in Fig. 20.30). An example of a low-frequency noise characterization system that we have used to study the noise in thin film polymer transistors is shown in Fig. 20.30. Note that LFN measurements are time intensive because a large number of averages are required for smooth spectra. Also, in noise measurements, the power spectrum densities S_V and S_I for the noise voltages and currents are measured, in units V^2/Hz and I^2/Hz , respectively.

20.6.2 Noise from the Interfacial Oxide Layer

Here we present some sample results and show how low-frequency noise spectra in ultrasmall devices can be used to estimate the oxide trap density. Generally, the low-

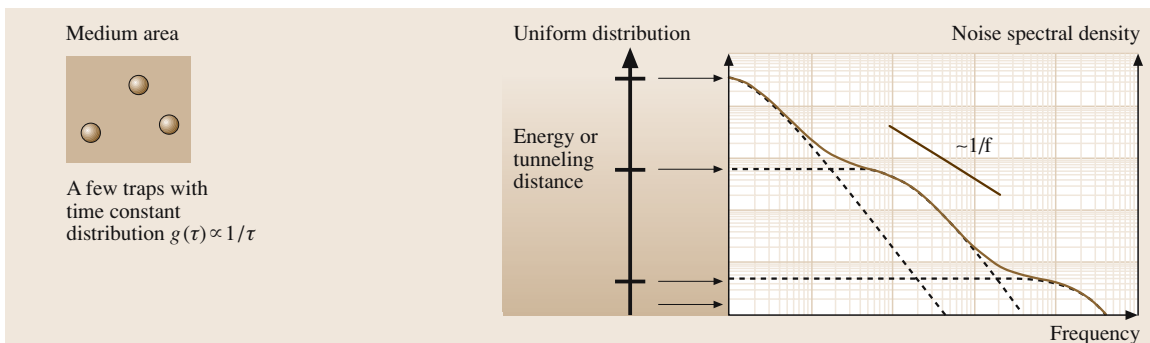


Fig. 20.32 Schematic representation of a “medium”-area PE-BJT ($\approx 0.5 \mu\text{m}^2$) with a few traps. Note that g - r bumps appear for each trap since there are only a few traps

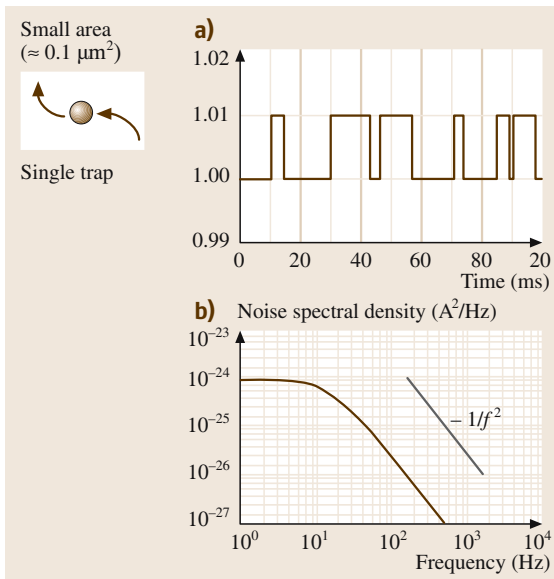


Fig. 20.33 Schematic representation of a “small”-area PE-BJT ($\approx 0.1 \mu\text{m}^2$) with one trap. Note that a single RTS and g-r spectrum appear because there is only one trap

base current, the noise spectra can be modeled as

$$S_{I_B} = \frac{K_F \cdot I_B^{A_F}}{f} + \sum_{i=1}^n \frac{B_i \tau_i}{1 + (2\pi f \tau_i)^2} + 2qI_B, \quad (20.83)$$

where the symbols have their usual meanings, see [20.68] for example.

As described in [20.63–68], the LFN in PE BJT originates from the thin layer of oxide between the monocrystalline and polycrystalline silicon emitter. The defects at this interface may be dangling oxygen bonds, oxygen vacancies, interface states or oxide traps [20.63]. Devices with large emitter areas have many traps, and these produce generation–recombination noise which produce $1/f$ noise when added. This is schematically shown in Fig. 20.31.

As the device area is reduced, and assuming a constant trap density (which is normally true for devices on the same wafer), then there are fewer traps in the interfacial oxide layer for smaller area devices. In this case, the spectral density of the noise changes and it gains characteristic “bumps” associated with resolvable g-r noise components. This is schematically shown in Fig. 20.32.

In very small devices with only a single trap, for example, the noise spectrum changes dramatically; only g-r noise is observed in the frequency domain along with a random telegraph signal (RTS) in the time domain. This is schematically shown in Fig. 20.33. Real experimental results are shown for three sizes of transistors (2.4 , 0.64 and $0.16 \mu\text{m}^2$) in Fig. 20.34. Here, one can see how $1/f$ noise is made up of g-r spectra as the emitter geometries are scaled to smaller and smaller values. For the PE BJT with an emitter area of $0.16 \mu\text{m}^2$, a lower bound of $\approx 10^9/\text{cm}^2$ can be approximated for the oxide trap density. Similar results have been obtained for MOSFETs [20.73].

20.6.3 Impedance Considerations During Noise Measurement

Two basic circuits can be employed when measuring the low-frequency noise (LFN) in a device. These configurations are sketched in Fig. 20.35. In voltage noise measurement (Fig. 20.35a), a low-noise preamplifier senses the voltage across the device, and this signal is sent to a spectrum analyzer or a fast Fourier transform (FFT) analyzer. In current noise measurements (Fig. 20.35b), the low-noise preamplifier senses the current through the device, converts it into a voltage, and forwards the voltage to a FFT or spectrum analyzer.

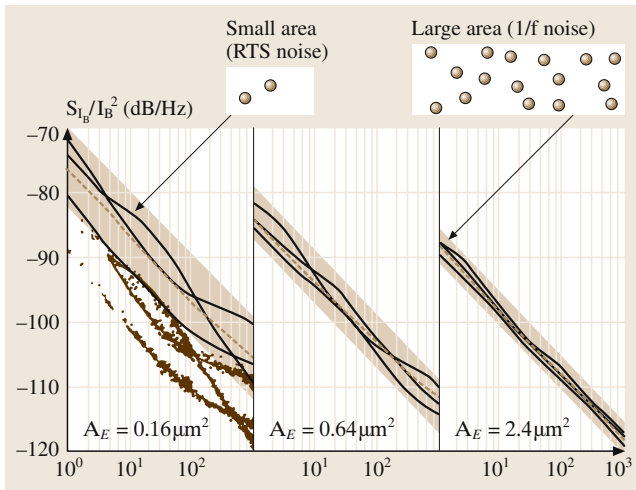


Fig. 20.34 Experimental results for low-frequency noise spectra from sets of large-, medium- and small-area PE-BJTs. In all cases, the average spectrum is $1/f$ noise and the relative magnitude of the $1/f$ noise is the same; that is, the area of $K_F \times$ is the same for the three sets of transistors. After [20.59–62]

frequency noise spectra of polysilicon emitter (PE) BJTs are made up of $1/f$ noise, generation–recombination (g-r) noise and shot noise sources. In the case of the

In principle, both configurations can be used for LFN measurement, but the impact of the nonideality of the amplifier (such as the input impedance, noise voltage and current) changes when the device impedance changes. Also, the noise from the bias source varies with each measurement set-up.

The noise equivalent circuit used for voltage measurement is shown in Fig. 20.36. The noise from the amplifier is represented by the input-referred noise voltage (S_{Vn}) and noise current (S_{In}) sources. The noise voltage from the bias is represented by S_{V0} . The impedance of the bias source is R_0 , whereas the input impedance of the amplifier is neglected, since it is usually very high compared to R_0 . The impedance of DUT is r_d . The noise current S_{Id} of the device that can be measured, assuming that the noise voltage at the input of the amplifier $S_{Vm} = S_{Id} \times r_d^2$. However, the amplifier sees a different level of S_{Vm} , given by

$$S_{Vm} = \frac{S_{OUT}}{A^2} = S_{V0} \left(\frac{r_d}{r_d + R_0} \right)^2 + S_{Vn} + (S_{In} + S_{Id})Z^2 \quad (20.84)$$

where

$$Z = (r_d // R_0) = \frac{r_d \cdot R_0}{r_d + R_0} \quad (20.85)$$

and A is the voltage gain of the amplifier. Therefore, the estimated value for S_{Id} is

$$S_{Id} = \frac{S_{Vm} - S_{Vn}}{Z^2} - \frac{S_{V0}}{R_0^2} - S_{In}. \quad (20.86)$$

The uncertainty in (20.86) is

$$\frac{\Delta S_{Id}}{S_{Id}} = \frac{\Delta S_{Vn}}{S_{Vm}} + \frac{\Delta S_{V0}}{S_{Vm}} \left(\frac{Z}{R_0} \right)^2 + \frac{\Delta S_{In}}{S_{Vm}} Z^2 \quad (20.87)$$

where $\Delta S \leq S$ denotes the uncertainty in each noise source. As seen from (20.87), the impact of the bias source noise ΔS_{V0} and the input current noise ΔS_{In} can be reduced if the impedance of the measurement circuit Z is low and the ratio r_d/R_0 is kept much less than 1; in other words, the voltage noise measurement is more appropriate for low-impedance devices, such as diodes at forward biasing, and the noise floor of the measurement is limited by the input-referred voltage noise S_{Vn} of the amplifier.

For the other (dual) case, current noise measurement, the noise equivalent circuit is shown in Fig. 20.37. The corresponding equations for the measured noise current S_{Im} , Z , the device noise S_{Id} , and the uncertainty,

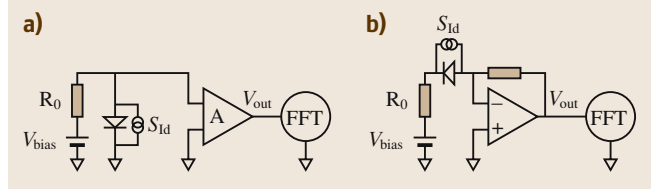


Fig. 20.35a,b Basic circuits used to measure the low-frequency noise (LFN) in a device (DUT). (a) Voltage noise measurement; (b) current noise measurement

respectively, are given by (20.88–20.91) below

$$S_{Im} = \frac{S_{OUT}}{R^2} = \frac{S_{V0}}{(r_d + R_0)^2} S_{In} + \frac{S_{Vn}}{Z^2} + S_{Id} \left(\frac{r_d}{r_d + R_0} \right)^2 \quad (20.88)$$

$$Z = (r_d + R_0) // R = \left(\frac{1}{R} + \frac{1}{r_d + R_0} \right)^{-1} \quad (20.89)$$

$$S_{Id} = (S_{Im} - S_{In}) \left(1 + \frac{R_0}{r_d} \right)^2 - \frac{S_{V0}}{r_d^2} - \frac{S_{Vn}}{Z^2} \left(1 + \frac{R_0}{r_d} \right)^2 \quad (20.90)$$

$$\frac{\Delta S_{Id}}{S_{Id}} = \frac{\Delta S_{In}}{S_{Im}} + \frac{\Delta S_{V0}}{S_{Im}} \frac{1}{(r_d + R_0)^2} + \frac{\Delta S_{Vn}}{Z^2 S_{Im}} \quad (20.91)$$

As expected from duality, it is apparent from (20.91) that the impact of the bias source noise ΔS_{V0} and the input

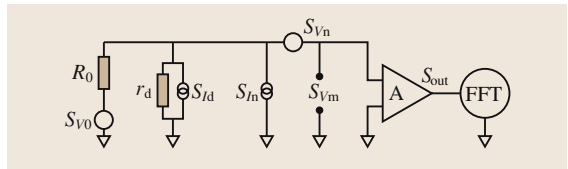


Fig. 20.36 Noise equivalent circuit for voltage noise measurements

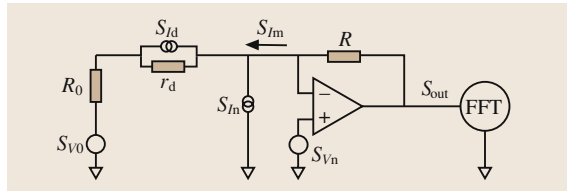


Fig. 20.37 Noise equivalent circuit of the current noise measurement

voltage noise ΔS_{In} can be reduced if the impedance of the measurement circuit Z and $(r_d + R_0)$ are both high; in other words, the *current noise measurement is more appropriate for high-impedance devices, such as diodes at reverse biasing*, and the noise floor of the measurement is limited by the input-referred current noise S_{I_n} of the amplifier.

This analysis above demonstrates that the choice of the measurement configuration follows our expectation that voltage should be measured in low-impedance devices and current in high-impedance devices. Also, the noise floor limiting parameter of the preamplifier is of the same type as the type of measurement;

that is, input-referred noise voltage for voltage noise measurement and input-referred noise current for current noise measurement. Note that there is a trade-off between the voltage and current noise in amplifiers, which implies that the measurement configuration – either voltage or current measurement – should also be carefully selected with respect to the impedance of the device under test. In addition, four-point connection can be used to measure the noise in very low impedance devices ($r_d < 100$). These and other considerations for low-frequency noise instrumentation are discussed in many papers, for example [20.74–76].

20.7 Deep-Level Transient Spectroscopy

Deep-level transient spectroscopy (DLTS) is a fairly complicated electrical characterization technique where the temperature is varied in large range from cryogenic temperatures (< 80 K) to well above room temperature (> 400 K). However, it is a powerful and versatile technique for investigating deep-level defects and it also gives accurate values for the capture cross-sections of defects. There are several DLTS techniques and [20.77, 78] provide recent reviews of the subject. In DLTS, the semiconductor device or junction is pulsed with an appropriate signal, and the resulting transient (such as capacitance, voltage or current) is monitored at different temperatures. Using these recorded transients at different temperatures, it is possible to generate a spectrum with peaks, each of which is associated with a deep level. The heights of the peaks are proportional to the defect density.

Here, we will focus on a new version of DLTS: the constant resistance (CR) DLTS technique [20.79–81]. We were able to accurately investigate bulk defects in a variety of test structures with CR-DLTS. Using body bias in a MOS transistor, we were able to distinguish interfacial and bulk defects that are important for different applications. For example, interfacial defects are important for electronic applications, and bulk defects are important for imaging or radiation detection applications. Examples of results from DLTS studies with and without body bias will be discussed.

CR-DLTS is well-suited to investigations of electrically active point defects that are responsible for the

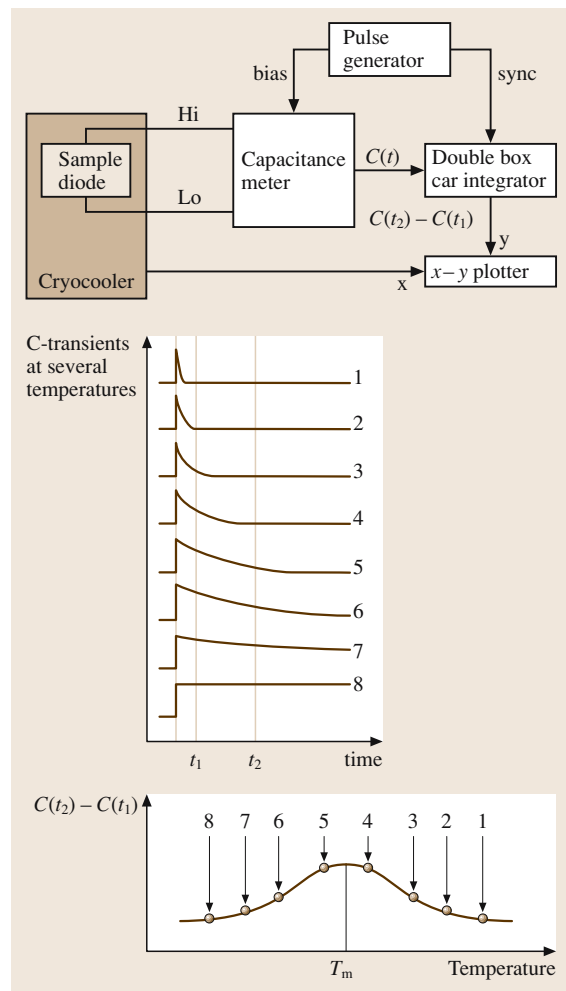


Fig. 20.38 Schematic representation of a conventional DLTS system. The time scans from which the DLTS temperature spectrum is obtained are shown on the right

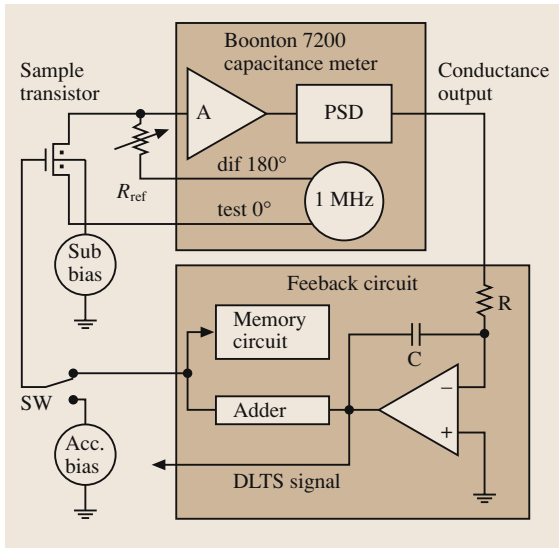


Fig. 20.39 Block diagram representation of the CR-DLTS system

creation of deep levels in the semiconductor band-gap. CR-DLTS can also be used to distinguish bulk traps and interface traps in MOSFETs.

A conventional DLTS system is shown schematically in Fig. 20.38. In DLTS, an excitation pulse is applied to the sample to fill all of the traps and then the pulsing is stopped. The next step is to detect the transient signal from the sample due to charge emission from the traps. The right side of Fig. 20.38 shows capacitance transients at eight different temperatures. By selecting a time window from t_1 to t_2 , and then plotting $[C(t_1) - C(t_2)]$ as a function of temperature, a DLTS spectrum with a characteristic peak is obtained as shown in the bottom of Fig. 20.38.

This peak is a signature from a specific defect level. To determine the properties of the defect (its energy level and capture cross-section), the time window ($\tau = t_2 - t_1$) is changed. In this case, different DLTS spectra are obtained at different temperatures. Using the time difference τ and the temperatures at which the peaks occur, Arrhenius plots are constructed in order to determine the defect energy level and its capture cross-section. Examples of DLTS spectra and Arrhenius plots associated with CR-DLTS are presented later (in Fig. 20.40).

A block representation of the CR-DLTS system is shown in Fig. 20.39. More details can be found in [20.79–83]. A discussion of the signal processing and averaging techniques used with this DLTS technique can be found in [20.83]. Here, the gate bias voltage of the

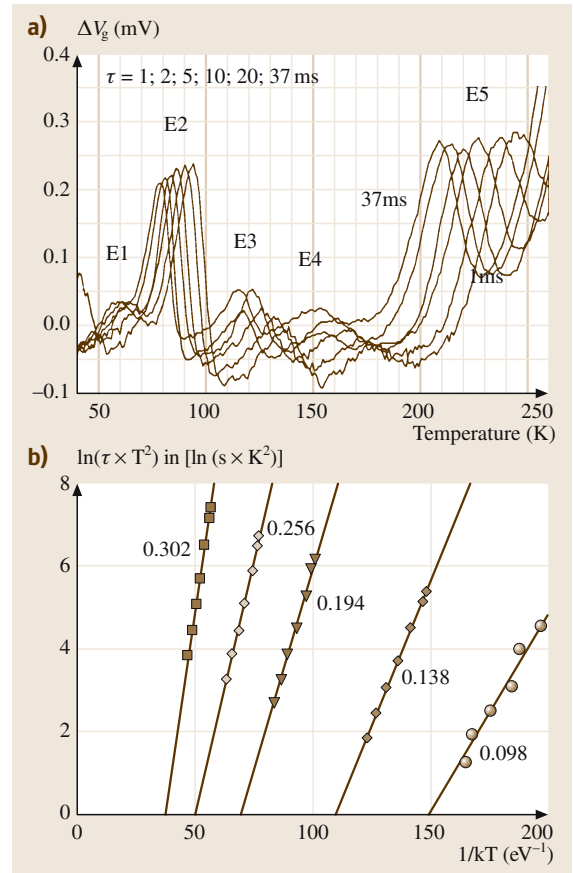


Fig. 20.40 (a) CR-DLTS spectra of a $50\ \mu\text{m} \times 20\ \mu\text{m}$ MOSFET damaged with 2.7×10^9 pronton/ cm^2 . (b) Arrhenius plot derived from the CR-DLTS spectra, showing the energies of the five traps E1–E5. After [20.80, 82]

field-effect transistor is adjusted using a feedback circuit so that the resistance corresponding to the source–drain conductance matches that of a reference resistor R_{ref} , which is typically around $1\ \text{M}\Omega$. The voltage transient due to the change in occupancy of the traps appears as a compensation voltage on the gate. This voltage change can be regarded as a threshold voltage change because the flat-band voltage of the device changes when the occupancy of the traps change. More details on how this change in the threshold is related to the traps can be found in [20.77, 79, 81].

Some important advantages of the CR DLTS technique are that the surface mobility of the MOS transistors does not need to be high, and that it is theoretically independent of the gate area of the transistor. This is expected, since the small amount of charge trapped beneath

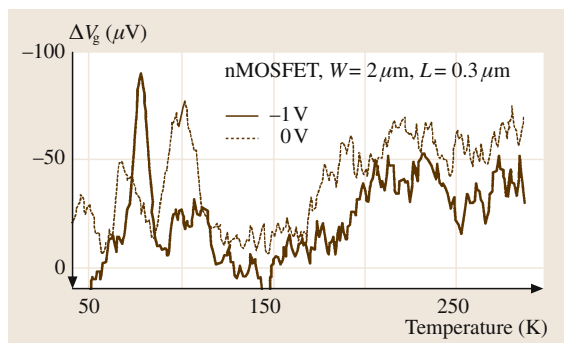


Fig. 20.41 Effect of body bias on CR-DLTS spectra. The body bias affects the surface and bulk traps in different ways

the gate must be balanced by a correction voltage applied across a relatively small gate–substrate capacitance.

Figure 20.40a shows six DLTS spectra for a JFET damaged with 2.7×10^9 protons/cm² [20.79, 81] with six selected rate windows. Using the temperatures at which the peaks occur and the rate windows, Arrhenius plots can be constructed as shown in

Fig. 20.40b, where the energies of five electron trap levels below the conduction band are also indicated. For the five traps, the extracted capture cross sections were 4.6×10^{-15} cm² (E1), 6.3×10^{-15} cm² (E2), 1.2×10^{-16} cm² (E3), 8.5×10^{-16} cm² (E4) and 3.4×10^{-15} cm² (E5).

Figure 20.41 shows CR-DLTS spectra as the source–body bias voltage is varied. The scans with a body bias of -1 V are lower in magnitude than those without substrate bias, except for the peak associated with the hole trap at 0.13 eV above the valence band [20.81]. When the reverse substrate bias is increased, the gate control of the space charge region near the channel decreases, meaning that fewer interface traps participate in the capture and emission of charges. However, the increased reverse substrate bias results in an increased space charge region in the silicon beneath the gate, so more bulk deep levels can participate in the capture and emission processes. This explains the increased deep-level peak (below 75 K) when -1 V is applied to the body. These differences between the CR-DLTS spectra demonstrate the ability to distinguish bulk traps from surface traps when the substrate bias of the MOSFET is varied.

References

- 20.1 M.J. Deen: Proc. Sixth Symp. Silicon Nitride and Silicon Dioxide Thin Insulating Films, ed. by R.E. Sah, M.J. Deen, D. Landheer, K.B. Sundaram, W.D. Brown, D. Misra (Electrochem. Soc. Paris, Paris 2003) pp. 3–21
- 20.2 P. Rai-Choudhury, J. Benton, D. Schroder (Eds.): Proc. Symp. Diagnostic Techniques Semiconductor Materials and Devices, Proc. Vol. 97–12 (The Electrochemical Society Press, New Jersey 1997)
- 20.3 D. Schroder: *Semiconductor Material and Device Characterization*, 2nd edn. (Wiley, New York 1998)
- 20.4 W. Runyan, T. Shaffner: *Semiconductor Measurements and Instrumentation*, 2nd edn. (McGraw Hill, New York 1997)
- 20.5 ITRS: *International Technology Roadmap for Semiconductors 2001 version*, URL: <http://public.itrs.net/Files/2001ITRS/Home.htm> (2001)
- 20.6 R. Pierret: *Advance Semiconductor Fundamentals*, Modular Ser. Solid State Dev. 6 (Addison-Wesley, Reading 1987)
- 20.7 W. Sawyer: Proc. 1998 Int. Conf. on Ion Implantation Technology, (IEEE Press, Piscataway, 1998)
- 20.8 L. J. van der Pauw: Philos. Res. Rev. **13**, 1–9 (1958)
- 20.9 M. Newsam, A. Walton, M. Fallon: Proc. 1996 Int. Conf. Microelectronic Test Structures (1996) pp. 247–252
- 20.10 T. Noda, D. Lee, H. Shim, M. Sakuraba, T. Matsuura, J. Murota: Thin Solid Films **380**, 57–60 (2000)
- 20.11 P. De Wolf, R. Stephenson, S. Biesemans, Ph. Jansen, G. Badenes, K. De Meyer, W. Vandervorst: Int. Electron Dev. Meeting (IEDM) Tech. Dig. (1998) pp. 559–562
- 20.12 J.H. Orchard-Webb, R. Coultier: Proc. IEEE Int. Conf. Microelectron. Test Structures, (1989) pp. 169–173
- 20.13 J. D. Plummer, P. B. Griffin: Proc. IEEE **89**(3) (2001) 240–258
- 20.14 S. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)
- 20.15 W. M. Loh, S. E. Swirhun, T. A. Schreyer, R. M. Swanson, K. C. Saraswat: IEEE Trans. Electron Dev. **34**(3), 512–524 (1987)
- 20.16 S. Zhang, M. Östling, H. Norström, T. Arnborg: IEEE Trans. Electron Dev. **41**(8), 1414–1420 (1994)
- 20.17 A. S. Holland, G. K. Reeves: Microelectron. Reliab. **40**, 965–971 (2000)
- 20.18 Y. Qiu: *Introduction to the Quantum Hall Effect*, URL: <http://www.pha.jhu.edu/~qiuy/qhe> (1997)
- 20.19 E. H. Hall: Am. J. Math. **2**, 287–292 (1879)
- 20.20 P. Elias, S. Hasenohrl, J. Fedor, V. Cambel: Sensors Actuat. A **101**, 150–155 (2002)
- 20.21 A. Vandooren, S. Cristoloveanu, D. Flandre, J. P. Colinge: Solid-State Electron. **45**, 1793–1798 (2001)

- 20.22 D. T. Lu, H. Ryssel: *Curr. Appl. Phys.* **1**(3–5), 389–391 (2001)
- 20.23 R. L. Petritz: *Phys. Rev.* **110**, 1254–1262 (1958)
- 20.24 P. Terziyska, C. Blanc, J. Pernot, H. Peyre, S. Conterras, G. Bastide, J. L. Robert, J. Camassel, E. Morvan, C. Dua, C. C. Brylinski: *Phys. Status Solidi A* **195**(1), 243–247 (2003)
- 20.25 G. Rutsch, R. P. Devaty, D. W. Langer, L. B. Rowland, W. J. Choyke: *Mat. Sci. Forum* **264–268**, 517–520 (1998)
- 20.26 P. Blood, J. W. Orton: *The Electrical Characterization of Semiconductor: Majority Carriers and Electron States*, (Techniques of Physics, Vol. 14) (Academic, New York 1992)
- 20.27 Q. Lu, M. R. Sardela Jr., T. R. Bramblett, J. E. Greene: *J. Appl. Phys.* **80**, 4458–4466 (1996)
- 20.28 S. Wagner, C. Berglund: *Rev. Sci. Instrum.* **43**(12), 1775–1777 (1972)
- 20.29 E. H. Nicollian, J. R. Brews: *MOS (Metal Oxide Semiconductor) Physics and Technology* (Wiley, New York 1982)
- 20.30 *Model 82–DOS Simultaneous C–V Instruction Manual* (Keithley Instruments, Cleveland 1988)
- 20.31 W. Beadle, J. Tsai, R. Plummer: *Quick Reference Manual for Silicon Integrated Circuit Technology* (Wiley, New York 1985)
- 20.32 J. Brews: *J. Appl. Phys.*, **44**(7), 3228–3231 (1973)
- 20.33 M. Kuhn: *Solid-State Electron.* **13**, 873–885 (1970)
- 20.34 C. N. Berglund: *IEEE Trans. Electron Dev.*, **13**(10), 701–705 (1966)
- 20.35 S. Witczak, J. Schuele, M. Gaitan: *Solid-State Electron.* **35**, 345 (1992)
- 20.36 M. J. Deen: *Electron. Lett.* **28**(3), 1195–1197 (1992)
- 20.37 Z. P. Zuo, M. J. Deen, J. Wang: *Proc. Canadian Conference on Electrical and Computer Engineering* (IEEE Press, Piscataway, 1989) pp. 1038–1041
- 20.38 A. Raychaudhuri, M. J. Deen, M. I. H. King, W. Kwan: *IEEE Trans. Electron Dev.* **43**(7), 1114–1122 (1996)
- 20.39 W. S. Kwan, A. Raychaudhuri, M. J. Deen: *Can. J. Phys.* **74**, S167–S171 (1996)
- 20.40 T. Matsuda, R. Takezawa, K. Arakawa, M. Yasuda, T. Ohzone, T. Kameda, E. Kameda: *Proc. International Conference on Microelectronic Test Structures (ICMTS 2001)* (IEEE Press, Piscataway, 2001) pp. 65–70
- 20.41 A. Raychaudhuri, M. J. Deen, M. I. H. King, W. Kwan: *IEEE Trans. Electron Dev.* **43**(1), 110–115 (1996)
- 20.42 G. Groeseneken, H. Maes, N. Beltram, R. DeKeersmaker: *IEEE Trans. Electron Dev.* **31**, 42–53 (1984)
- 20.43 X. Li, M. J. Deen: *Solid-State Electron.* **35**(8), 1059–1063 (1992)
- 20.44 X. M. Li, M. J. Deen: *IEEE International Electron Devices Meeting (IEDM)* (IEEE Press, Piscataway, 1990) pp. 85–87
- 20.45 D. S. Ang, C. H. Ling: *IEEE Electron Dev. Lett.* **19**(1), 23–25 (1998)
- 20.46 H. Uchida, K. Fukuda, H. Tanaka, N. Hirashita: *International Electron Devices Meeting* (1995) pp. 41–44
- 20.47 N. S. Saks, M. G. Ancona: *IEEE Trans. Electron Dev.* **37**(4), 1057–1063 (1990)
- 20.48 S. Mahapatra, C. D. Parikh, V. R. Rao, C. R. Viswanathan, J. Vasi: *IEEE Trans. Electron Dev.* **47**(4), 789–796 (2000)
- 20.49 Y.-L. Chu, D.-W. Lin, C.-Y. Wu: *IEEE Trans. Electron Dev.* **47**(2), 348–353 (2000)
- 20.50 J. S. Bruglers, P. G. Jespers: *IEEE Trans. Electron Dev.* **16**, 297 (1969)
- 20.51 D. Bauza: *J. Appl. Phys.* **94**(5), 3239–3248 (2003)
- 20.52 L. M. Head, B. Le, T. M. Chen, L. Swiatkowski: *Proceedings 30th Annual International Reliability Physics Symposium* (1992) pp. 228–231
- 20.53 S. An, M. J. Deen: *IEEE Trans. Electron Dev.* **47**(3), 537–543 (2000)
- 20.54 S. An, M. J. Deen, A. S. Vetter, W. R. Clark, J.-P. Noel, F. R. Shepherd: *IEEE J. Quantum Elect.* **35**(8), 1196–1202 (1999)
- 20.55 M. J. Deen, C. Quon: *7th Biennial European Conference – Insulating Films on Semiconductors (INFOS 91)*, ed. by W. Eccleston, M. Uren (IOP Publishing Ltd., Liverpool U.K., 1991) pp. 295–298
- 20.56 Z. Celik-Butler: *IEE P.-Circ. Dev. Syst.* **149**(1), 23–32 (2002)
- 20.57 J. Chen, A. Lee, P. Fang, R. Solomon, T. Chan, P. Ko, C. Hu: *Proceedings IEEE International SOI Conference* (1991) pp. 100–101
- 20.58 J. Sikula: *Proceedings of the 17th International Conference on Noise in Physical Systems and 1/f Fluctuations (ICNF 2003)*, (CNRL, Prague, 2003)
- 20.59 M. J. Deen, Z. Celik-Butler, M. E. Levinhstein (Eds.): *SPIE Proc. Noise Dev. Circ.* **5113** (2003)
- 20.60 C. R. Doering, L. B. Kish, M. Shlesinger: *Proceedings of the First International Conference on Unsolved Problems of Noise*, (World Scientific Publishing, Singapore, 1997)
- 20.61 D. Abbot, L. B. Kish: *Proceedings of the Second International Conference on Unsolved Problems of Noise and Fluctuations*, (American Institute of Physics Conference Proceedings: 511, Melville, New York 2000)
- 20.62 S. M. Bezrukov: *Proceedings of the Third International Conference on Unsolved Problems of Noise and Fluctuations*, Washington, DC (American Institute of Physics Conference Proceedings: 665, Melville, New York 2000)
- 20.63 M. J. Deen, S. L. Rumyantsev, M. Schroter: *J. Appl. Phys.* **85**(2), 1192–1195 (1999)
- 20.64 M. Sanden, O. Marinov, M. Jamal Deen, M. Ostling: *IEEE Electron Dev. Lett.*, **22**(5), 242–244 (2001)
- 20.65 M. Sanden, O. Marinov, M. Jamal Deen, M. Ostling: *IEEE Trans. Electron Dev.* **49**(3), 514–520 (2002)
- 20.66 M. J. Deen, J. I. Ilowski, P. Yang: *J. Appl. Phys.* **77**(12), 6278–6288 (1995)

- 20.67 M.J. Deen, E. Simoen: IEE P.-Circ. Dev. Syst. **49**(1), 40–50 (2002)
- 20.68 M.J. Deen: IEE Proceedings – Circuits, Devices and Systems – Special Issue on Noise in Devices and Circuits **151**(2) (2004)
- 20.69 M.J. Deen, O. Marinov: IEEE Trans. Electron Dev. **49**(3), 409–414 (2002)
- 20.70 O. Marinov, M.J. Deen, J. Yu, G. Vamvoudnis, S. Holdcroft, W. Woods: Instability of the Noise Level in Polymer Field Effect Transistors with Non-Stationary Electrical Characteristics, Third International Conference on Unsolved Problems of Noise and Fluctuations (UPON 02), Washington, DC (AIP Press, Melville, 2002)
- 20.71 M. Marin, M.J. Deen, M. de Murcia, P. Llinares, J. C. Vildeuil: IEE P.-Circ. Dev. Syst. **151**(2), 95–101 (2004)
- 20.72 A. Chandrakasan: Proceedings European Solid-State Circuits Conference (ESSCIRC 2002), (AIP Press, Melville, 2002) pp. 47–54
- 20.73 R. Brederlow, W. Weber, D. Schmitt-Landsiedel, R. Thewes: IEDM Technical Digest (1999) pp. 159–162
- 20.74 L. Chaar, A. van Rheenen: IEEE Trans. Instrum. Meas. **43**, 658–660 (1994)
- 20.75 C.-Y. Chen, C.-H. Kuan: IEEE Trans. Instrum. Meas. **49**, 77–82 (2000)
- 20.76 C. Ciofi, F. Crupi, C. Pace, G. Scandurra: IEEE Trans. Instrum. Meas. **52**, 1533–1536 (2003)
- 20.77 P. Kolev, M.J. Deen: Development and Applications of a New DLTS Method and New Averaging Techniques. In: *Adv. Imag. Electr. Phys.*, ed. by P. Hawkes (Academic, New York 1999)
- 20.78 P. McLarty: Deep Level Transient Spectroscopy (DLTS). In: *Characterization Methods for Submicron MOSFETs*, ed. by H. Haddara (Kluwer, Boston 1996) pp. 109–126
- 20.79 P. V. Kolev, M. J. Deen: J. Appl. Phys. **83**(2), 820–825 (1998)
- 20.80 P. Kolev, M.J. Deen, T. Hardy, R. Murowinski: J. Electrochem. Soc. **145**(9), 3258–3264 (1998)
- 20.81 P. Kolev, M.J. Deen, J. Kierstead, M. Citterio: IEEE Trans. Electron Dev. **46**(1), 204–213 (1999)
- 20.82 P. Kolev, M.J. Deen: Proceedings of the Fourth Symposium on Low Temperature Electronics and High Temperature Superconductivity, **97**–2, ed. by C. Claeys, S.I. Raider, M.J. Deen, W.D. Brown, R. K. Kirschman (The Electrochemical Society Press, New Jersey, 1997) pp. 147–158
- 20.83 P. V. Kolev, M.J. Deen, N. Alberding: Rev. Sci. Instrum. **69**(6), 2464–2474 (1998)

Materials

Part C

Part C Materials for Electronics

**21 Single-Crystal Silicon:
Electrical and Optical Properties**

Shlomo Hava, Beer Sheva, Israel
Mark Auslender, Beer Sheva, Israel

**22 Silicon–Germanium:
Properties, Growth and Applications**

Peter Ashburn, Southampton, UK
Darren M. Bagnall, Southampton, UK

23 Gallium Arsenide

Mike Brozel, Glasgow, UK

**24 High-Temperature Electronic Materials:
Silicon Carbide and Diamond**

Magnus Willander, Göteborg, Sweden
Milan Friesel, Göteborg, Sweden
Qamar-ul Wahab, Linköping, Sweden
Boris Straumal, Chernogolovka, Russia

**25 Amorphous Semiconductors: Structure,
Optical, and Electrical Properties**

Kazuo Morigaki, Tokyo, Japan
Chisato Ogihara, Ube, Japan

26 Amorphous and Microcrystalline Silicon

Akihisa Matsuda, Chiba, Japan

27 Ferroelectric Materials

Roger Whatmore, Lee Maltings, Ireland

28 Dielectric Materials for Microelectronics

Robert M. Wallace, Richardson, USA

29 Thin Films

Robert D. Gould[†], Keele, UK

30 Thick Films

Neil White, Highfield, UK

Single-Crystal Silicon

21. Single-Crystal Silicon: Electrical and Optical Properties

Electrical and optical properties of crystalline semiconductors are important parts of pure physics and material science research. In addition, knowledge of parameters related to these properties, primarily for silicon and III–V semiconductors, has received a high priority in microelectronics and optoelectronics since the establishment of these industries. For control protocols, emphasis has recently been placed on novel optical measurement techniques, which have proved very promising as nondestructive and even non-contact methods. Earlier they required knowledge of the free-carrier-derived optical constants, related to the electrical conductivity at infrared frequencies, but interest in the optical constants of silicon in the visible, ultraviolet (UV) and soft-X-ray ranges has been revived since the critical dimensions in devices have become smaller.

This chapter surveys the electrical (Sect. 21.2) and optical (Sect. 21.3) properties of crystalline silicon. Section 21.2 overviews the basic concepts. Though this section is bulky and its material is documented in textbooks, it seems worth including since the consideration here focuses primarily on silicon and is not spread over other semiconductors – this makes the present review self-contained. To avoid repeated citations we, in advance, refer the reader to stable courses on solid-state physics (e.g. [21.1, 2]), semiconductor physics (e.g. [21.3]), semiconductor optics (e.g. [21.4]) and electronic devices (e.g. [21.5]); seminal papers are cited throughout Sect. 21.2.

21.1	Silicon Basics	441
21.1.1	Structure and Energy Bands.....	441
21.1.2	Impurity Levels and Charge-Carrier Population ...	443
21.1.3	Carrier Concentration, Electrical and Optical Properties	446
21.1.4	Theory of Electrical and Optical Properties	447
21.2	Electrical Properties	451
21.2.1	Ohm's Law Regime	451
21.2.2	High-Electric-Field Effects.....	465
21.2.3	Review Material	471
21.3	Optical Properties	472
21.3.1	Diversity of Silicon as an Optical Material	472
21.3.2	Measurements of Optical Constants.....	472
21.3.3	Modeling of Optical Constants.....	474
21.3.4	Electric-Field and Temperature Effects on Optical Constants.....	477
	References	478

We realize how formidable our task is – publications on electrical and optical properties of silicon amount to a huge number of titles, most dating back to the 1980s and 1990s – so any review of this subject will inevitably be incomplete. Nevertheless, we hope that our work will serve as a useful shortcut into the silicon world for a wide audience of applied physics, electrical and optical engineering students.

21.1 Silicon Basics

21.1.1 Structure and Energy Bands

Normally silicon (Si) crystallizes in a diamond structure on a face-centered cubic (f.c.c.) lattice, with a lattice constant of $a_0 = 5.43 \text{ \AA}$. The basis of the diamond structure consists of two atoms with coordinates (0, 0, 0) and $a_0/4(1, 1, 1)$, as seen in Fig. 21.1. Other solids that can

crystallize in the diamond structure are C, Ge and Sn. The important notion for the electronic band structure is the Brillouin zone (BZ). The BZ is a primitive cell in the reciprocal-space lattice, which proves to be a body-centered cubic (b.c.c.) lattice for an f.c.c. real-space lattice. For this case, the BZ with important reference points and directions within it is shown in Fig. 21.2.

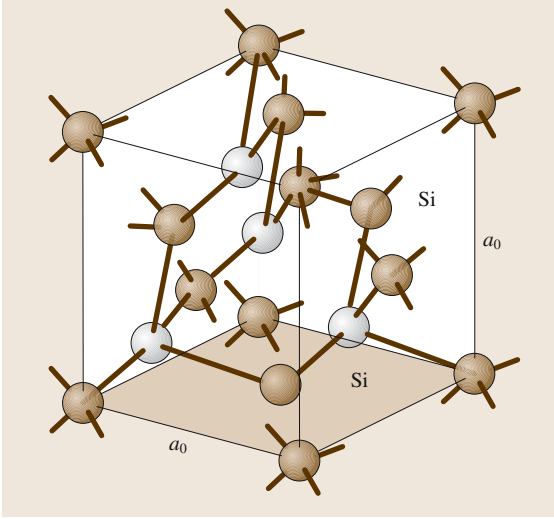


Fig. 21.1 Diamond crystal structure of Si

The states of electrons in solids are described by wave functions of the Bloch type

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{s\mathbf{k}}(\mathbf{r}), \quad (21.1)$$

where \mathbf{k} is the wave vector that runs over reciprocal space, s is a band index and $u_{s\mathbf{k}}(\mathbf{r})$ is the periodic function of the direct lattice (Bloch amplitude). Both $u_{s\mathbf{k}}(\mathbf{r})$ and the corresponding energy-band spectrum $E_s(\mathbf{k})$ are periodic in \mathbf{k} , which allows one to restrict consideration

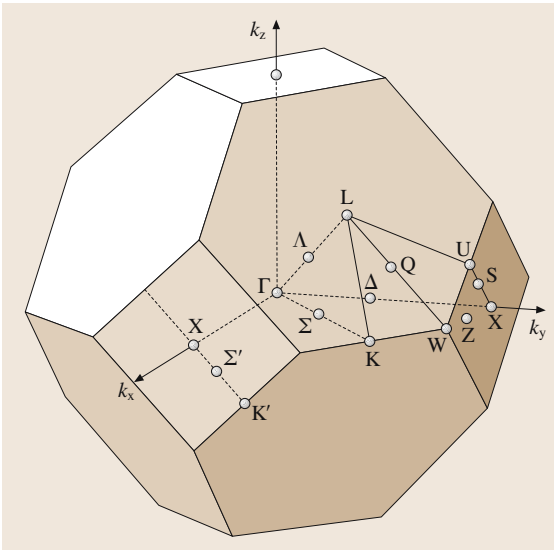


Fig. 21.2 Brillouin zone of the f.c.c. lattice

to within the BZ. The bands are arranged so that there are energy regions for which no states given by (21.1) exist. Such forbidden regions are called energy gaps or band gaps and result from the interaction of valence electron with the ion cores of crystal. In semiconductor science the term *band gap* is accepted for the energy distance between the maximum of $E_s(\mathbf{k})$ for the highest filled (valence) band and the minimum of $E_s(\mathbf{k})$ for the lowest empty (conduction) band (denoted by E_g). The band gap is called *direct* if the aforementioned maximum and minimum occur at the same point of the BZ, e.g. Γ (Fig. 21.2), and *indirect* if they occur at different points of the BZ, e.g. Γ and X (Fig. 21.2).

Si is an indirect-band-gap semiconductor with $E_g = 1.1700$ eV at 4.2 K. The calculated energy-band structure, that is the curves of $E_v(\mathbf{k})$ for selected directions in the BZ, is shown in Fig. 21.3a. The conduction-band minimum lies at six equivalent points Δ on the Γ -X lines (Fig. 21.2). In some vicinity (called the *valley*) of every such point the band spectrum is quadratic in \mathbf{k} , e.g. for the valley $\langle 100 \rangle$

$$E_c(\mathbf{k}) = E_{c0} + \frac{\hbar^2 (k_x - k_0)^2}{2m_l} + \frac{\hbar^2 (k_y^2 + k_z^2)}{2m_t}, \quad (21.2)$$

where $k_0 \approx 1.72\pi/a_0$, m_l and m_t are the longitudinal and transverse effective masses. The spectra for other five valleys are obtained from (21.2) by 90° rotations and inversions $k_0 \rightarrow -k_0$. Though the constant-energy surface for (21.2) is an ellipsoid (Fig. 21.3b), the density of states (DOS) proves to be the same as for an isotropic parabolic spectrum with an effective mass

$$m_{de} = 6^{2/3} m_l^{1/3} m_t^{2/3}, \quad (21.3)$$

which is called the *DOS effective mass*. Another mass, m_{ce} , which appears in the direct-current (DC) and optical conductivity formulas, is defined via the harmonic mean

$$\frac{1}{m_{ce}} = \frac{1}{3} \left(\frac{1}{m_l} + \frac{2}{m_t} \right). \quad (21.4)$$

For this reason, m_c is called the *conductivity/optical effective mass*. Equation (21.2) holds at $E_c - E_{c0} < 0.15$ eV, but at larger energies the ellipsoids warp strongly, especially near the X point; the change of the spectrum with energy is mostly due to the increasing m_t , while m_l increases weakly [21.6].

The valence-band maximum is at the Γ point ($\mathbf{k} = 0$), where the Bloch-wave state $u_{n0}(\mathbf{r})$ has the full symmetry of an atomic p-orbital, being six-fold degenerate in the nonrelativistic limit. The spin-orbit interaction splits

off from the bare band top a four-fold degenerate $p_{3/2}$ level up by $1/3\Delta_{so}$, and a two-fold degenerate $p_{1/2}$ level down by $-2/3\Delta_{so}$. The spectrum at $\mathbf{k} \neq 0$, even at energies near the top E_{v0} , is very complex. It consists of three branches, which are in general nonparabolic and nonisotropic [21.7]. At $E_{v0} - E_v \ll \Delta_{so}$ there are two nonparabolic anisotropic $p_{3/2}$ -derived sub-bands with the energy spectra

$$E_{v1,2}(\mathbf{k}) = E_{v0} + \frac{\hbar^2}{2m_0} \left[Ak^2 \pm \sqrt{B^2k^4 + C^2(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2)} \right],$$

$$A < 0 \quad (21.5)$$

and at $E_{v0} - E_v \ll 2\Delta_{so}$ there is an isotropic parabolic $p_{1/2}$ -derived split-off band with the energy spectrum

$$E_{v3}(\mathbf{k}) = E_{v0} - \Delta_{so} + \frac{\hbar^2k^2}{2m_0} A',$$

$$A' < 0. \quad (21.6)$$

Here m_0 is the free electron mass, A , B , C and A' are inverse hole-mass parameters (for small Δ_{so} , $A' \approx A$). In (21.5) the $+$ sign corresponds to heavy holes and the $-$ sign to light holes. The constant-energy surfaces for (21.5) are warped spheres (Fig. 21.3c), the DOS is nevertheless parabolic and is described by the effective mass

$$m_{dh} = \left(m_{d1}^{3/2} + m_{d2}^{3/2} \right)^{2/3}, \quad (21.7)$$

where $m_{d1,2}$ are the partial DOS effective masses. Quite different masses enter various physical quantities for heavy (light) holes, and are complicated functions of A , B and C ([21.9]); however, the split-off band is characterized by only one effective mass: $m_3 = m_0/|A'|$.

These band-structure parameters, obtained from cyclotron-resonance data and calculations, are presented in Tables 21.2 and 21.3. The unreferenced values of m_{de} , m_{ce} and m_3 were calculated from the referenced data using (21.3, 4) and the assumption $A' = A$. Optical data for m_{ce} are discussed in Sect. 21.3, while the specific-heat data for m_{de} are not considered here.

Considerable uncertainty and errors in the values of B and C have a small effect on the light-hole effective-mass values (m_{d2} and optical m_{c2}) but lead to an ambiguity in the heavy-hole effective-mass (m_{d1} and optical m_{c1}) values (Table 21.3). In each data set $m_{d2} \approx m_{c2}$, which means that an isotropic approximation is reasonable for the light-hole spectrum.

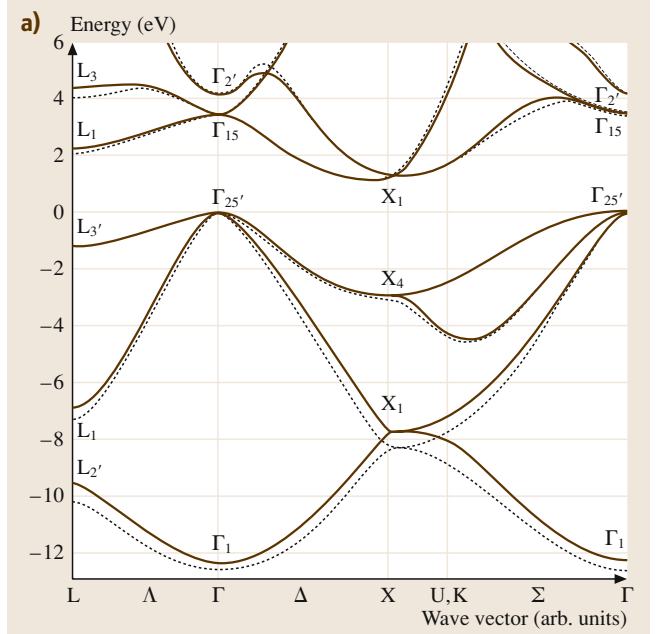


Fig. 21.3 Electronic band structure of Si: (a) Energy dispersion curves near the fundamental gap. After [21.8] with permission;

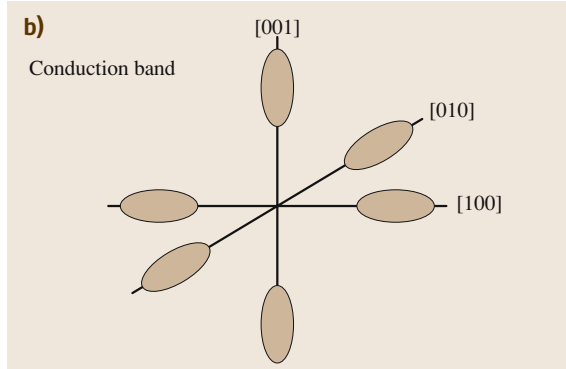
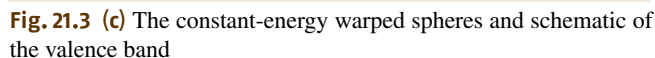


Fig. 21.3 (b) The constant-energy ellipsoids of the conduction band;

The experimental data [21.10, 11] and band-structure calculations [21.12] are in good agreement; the former is used in theoretical papers on hole transport in Si [21.13, 14].

Nonparabolic parts in the electron [21.6] and hole spectra [21.7] lead to apparent dependence of effective-mass measurements on the temperature and carrier concentration [21.15]. Recently the DOS mass issue was revisited [21.16] in connection with the intrinsic carrier concentration.



In highly purified Si electrons and holes are thermally generated with equal concentrations ($n = p$), equal to the intrinsic concentration n_i given by

where k_B and h are the Boltzmann and Planck constants, respectively, and T is the absolute temperature. To assess $n_i(T)$ theoretically one should take into account possible changes of $m_{de,h}$ and E_g with T . An early experimental plot of $n_i(T)$ for Si is shown in Fig. 21.4, giving $n_i(300\text{ K}) = 1.38 \times 10^{10}\text{ cm}^{-3}$. This and the early textbook value ($1.45 \times 10^{10}\text{ cm}^{-3}$) do not fit (21.8) with the established low-temperature values of $m_{de,h}$ and $E_g(300\text{ K}) = 1.1242\text{ eV}$, giving $(0.662\text{--}0.694) \times 10^{10}\text{ cm}^{-3}$ for the best available data.

In impure samples $n \neq p$; the charge-neutrality condition is then

together with the known expressions for n , p (with $np = n_i^2$ holding until the carriers become degenerate) and the concentrations of charged donors N_d^+ and acceptors N_a^- via the Fermi energy E_F and impurity ionization energies

Table 21.1 Energy levels of impurities in Si

Impurity	Donor/ acceptor	Below conduction band (eV)	Above conduction band (eV)
Ag	D	0.310	
	A		0.210
Al	A		0.057
As	D	0.049	
Au	D		0.330
	A	0.540	
B	A		0.045
Cu	A		0.490
	D		0.240
Fe	D	0.550	
	D		0.400
Ga	A		0.065
In	A		0.160
Li	D	0.033	
Mn	D	0.530	
P	D	0.044	
Sb	D	0.039	
S	D	0.180	
	D	0.370	
Zn	A	0.550	
	A		0.300

Table 21.2 The conduction-band effective masses and valence-band parameters for Si

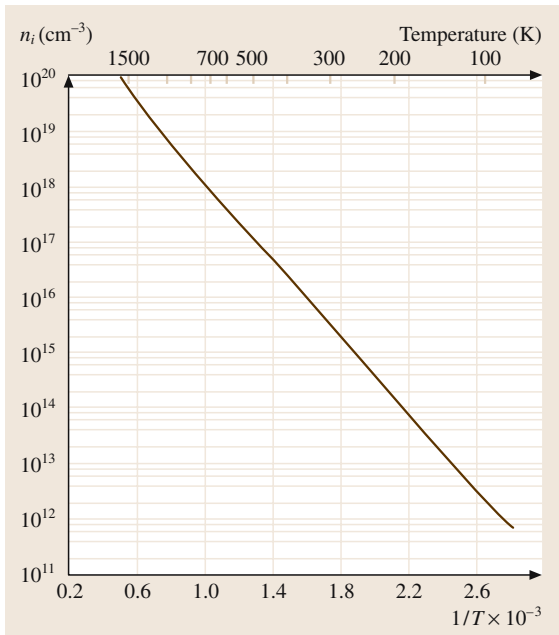
Conduction band					
Mass	m_l/m_0	m_t/m_0	m_{de}/m_0	m_{ce}/m_0	
Experiment	0.97 ± 0.04^a	0.19 ± 0.01^a	1.08 ± 0.05	0.26 ± 0.01	
	0.9163 ± 0.04^b	0.1905 ± 0.0001^b	1.0618 ± 0.0005	0.2588 ± 0.0001	
Theory	0.971^c	0.205^c	1.137	0.2652	
	0.9716^d	0.1945^d	1.0978	0.2780	
Valence band					
Parameter	Δ_{so}	A	$ B $ or B	$ C $	m_3/m_0
Experiment	0.0441 ± 0.004^g	-4.1 ± 0.2^a	1.6 ± 0.2^a	3.3 ± 0.2^a	0.24 ± 0.01
		-4.28 ± 0.02^e	-0.75 ± 0.04^e	5.25 ± 0.05^e	0.234 ± 0.001
		-4.27 ± 0.02^h	-0.63 ± 0.08^h	4.93 ± 0.15^h	0.234 ± 0.001
Theory		-4.38^c	0.84^c	4.11^c	0.23
	0.04^d	-4.38^d	-1.00^d	4.80^d	0.23
	0.044^f	-4.22^f	-0.78^f	4.80^f	0.24

^a [21.19], ^b [21.20], ^c [21.21], ^d [21.22], ^e [21.10], ^f [21.12], ^g [21.23], ^h [21.11]

Table 21.3 The valence-band effective masses calculated using experimental data

Mass	m_{d1}/m_0	m_{d2}/m_0	m_d/m_0	m_{c1}/m_0	m_{c2}/m_0
Exp. ^a	0.55 ± 0.12	0.16 ± 0.01	0.61 ± 0.12	0.51 ± 0.10	0.16 ± 0.01
Exp. ^b	0.58 ± 0.02	0.151 ± 0.001	0.63 ± 0.02	0.43 ± 0.01	0.145 ± 0.001
Exp. ^c	$0.49\text{--}0.56$	$0.153\text{--}0.158$	$0.54\text{--}0.62$	$0.40\text{--}0.43$	$0.147\text{--}0.152$

^a [21.19], ^b [21.10], ^c [21.11]

**Fig. 21.4** Intrinsic concentration in Si versus temperature. After [21.24] with permission

E_d and E_a allows one to calculate E_F (and hence all the concentrations) as a function of T and doping.

An impurity is called *shallow* if $|E_{c(v)0} - E_{d(a)}| \ll E_g$, and *deep* if $|E_{c(v)0} - E_{d(a)}| \approx 0.5E_g$. Shallow group V donors (Sb, P, As) and group III acceptors (B, Al, Ga, In) are well soluble in Si [21.25]. The ionization energy was calculated using the effective-mass approximation, and the Bohr model for donors [21.26–28] (Kohn and Luttinger, 1955; Kohn, 1955) and acceptors (Luttinger and Kohn, 1955). The value of $E_{d(a)}$ calculated in this way is insensitive to the specific shallow donor (acceptor). Actual thermal (i.e. retrieved from electrical measurements) ionization energies are given in Table 21.1. It is seen that shallow impurities have different ionization energies. This difference is small for donors and larger for acceptors. The same trend was observed for optical ionization energies, although they are different from the thermal values [21.29,30]. For shallow impurities, $E_{d(a)}$ decreases as doping becomes heavier, and becomes zero as $N_{d(a)}$ approaches the corresponding insulator–metal transition concentration [21.31].

Deep impurities (except for Mn, Fe and Zn) are amphoteric, i.e. they act simultaneously as donors and acceptors. For such a donor (acceptor) state, the level

may lie closer to the valence (conduction) than the conduction (valence) band. The deep impurities are mostly unionized at room temperature due to their large $E_{d(a)}$, so their direct contribution to n or p is negligible. The unionized deep impurities may, however, trap the carriers available from the shallow impurities or injection, thus decreasing the conductivity or the minority-carrier lifetime. Atoms that behave in Si in this manner, for example Au, Ag and Cu, are added for lifetime control. The properties of these impurities in Si have been studied in detail (see, e.g., [21.32, 33]).

21.1.3 Carrier Concentration, Electrical and Optical Properties

Concentration and Electrical Measurements

Measurements of carrier concentrations, as well as electrical and optical characteristics are most tractable if either $n \gg p$ (strongly n-type conduction) or $p \gg n$ (strongly p-type conduction). Since the np product is constant versus doping, the contribution of minority carriers to the conductivity becomes unimportant when $N_{d(a)}$ increases significantly over n_i . A standard route for determining $n(p)$ is Hall-effect measurements. The Hall coefficient R^H , measured directly on a long thin slab in a standard crossed electric and magnetic field configuration, is retrieved by

$$R^H = 10^{-8} \frac{V_H d}{IB}, \quad (21.10)$$

where V_H is the Hall voltage (Volts), I is the current (Amps), d is the sample thickness (cm) in the z -direction, and B is the magnetic field strength (Gauss) applied in this direction. There are two limiting cases. One, the high-field regime is defined by $qB\tau/mc \gg 1$, where m and τ are the appropriate mass and relaxation-time parameters, respectively. In this case

$$\begin{aligned} R_e^H(\infty) &= -\frac{1}{qn}, \\ R_h^H(\infty) &= \frac{1}{qp}. \end{aligned} \quad (21.11)$$

The other, low-field, regime holds with the opposite inequality; in this regime

$$\begin{aligned} R_e^H(0) &= -\frac{r_e}{qn}, \\ R_h^H(0) &= \frac{r_h}{qp}, \end{aligned} \quad (21.12)$$

where the constant of proportionality $r_{e(h)}$, called the electron (hole) Hall factor, depends on the details of the

scattering process and band structure. Thus the majority-carrier concentration is determined directly from $R_{e(h)}^H$ using a high-field Hall measurement. For typical laboratory magnetic fields, this regime is attainable only with extremely high mobility and low effective mass, which excludes moderately and heavily doped Si, for which very high magnetic fields are required. In some cases the Hall factor is quite close to unity, e.g. $r_e = 3\pi/8$ for the phonon scattering in the isotropic and parabolic (standard) band.

The electrical properties are fully described by the drift-diffusion relation for the electron (hole) current density $j_{e(h)}$

$$\begin{aligned} j_e &= -qn v_{de} + qD_e \nabla n, \\ j_h &= qp v_{dh} - qD_h \nabla p, \end{aligned} \quad (21.13)$$

where $v_{de} = -\mu_e E$ ($v_{dh} = \mu_h E$) is the drift velocity, $\mu_{e(h)}$ is the drift mobility, E is the electric field strength, and $D_{e(h)}$ is the diffusion coefficient; in general, $\mu_{e(h)}$ and $D_{e(h)}$ depend on E . In the homogeneous case (21.13) converts into the material equation $j_{e(h)} = \sigma_{e(h)} E$, where $\sigma_{e(h)} = qn\mu_{e(h)}$ is the electron (hole) conductivity; the total conductivity equals $\sigma = \sigma_e + \sigma_h$. In the weak-field DC (Ohm) and alternating current (AC: microwave or light, except for intense laser, irradiation) regimes, $D_{e(h)}$ is proportional to $\mu_{e(h)}$ being both constant versus E , depending on the radiation frequency ω .

Combining the high-induction Hall and Ohm resistivity ($\rho = \sigma^{-1}$) measurements one obtains the drift mobility

$$\mu_{e(h)} = R_{e(h)}^H(\infty) \sigma. \quad (21.14)$$

Replacing $R_e^H(\infty)$ by $R_e^H(0)$ in the right-hand side of (21.14), one arrives at the so-called Hall mobility

$$\mu_{e(h)}^H = R_{e(h)}^H(0) \sigma = r_{e(h)} \mu_{e(h)}, \quad (21.15)$$

which never equals the drift mobility, although it may be fairly close to it in the cases mentioned above. In general, to extract $n(p)$ from $R_e^H(0)$, the calculation of the $r_{e(h)}$ factor is completed. Magnetoresistance (MR), i.e. ρ versus B measurement, is an important experimental tool as well. Another established method is the Haynes–Shockley experiment, which allows one to measure the minority-carrier drift mobility. In high-electric-field conditions, a noise-measurement technique is used. A relatively novel, time-of-flight technique was used in the latest (to our knowledge) mobility and diffusion-coefficient measurements on lightly doped crystalline Si samples, both in the low- and high-field regimes [21.14].

Basic Optical Parameters

The electromagnetic response of homogeneous nonmagnetic material is governed by the dielectric constant tensor ε , which connects the electric displacement vector \mathbf{D} inside the material to \mathbf{E} through the material equation $\mathbf{D} = \varepsilon \mathbf{E}$. For cubic crystals, such as Si, ε is a scalar. An effective-medium homogeneous dielectric constant may be attributed to inhomogeneous and composite materials if the nonhomogeneity feature size is smaller than the radiation wavelength $\lambda = 2\pi c/\omega$. Actually, ε characterizes the material's bulk and therefore loses its sense in nanoscale structures (superlattices, quantum wells etc). The dependence $\varepsilon(\omega)$ expresses the optical dispersion in the material. The dielectric constant is usually represented via its real and imaginary parts: $\varepsilon = \varepsilon_1 + i\varepsilon_2$ ($\varepsilon_2 \geq 0$), connected to each other by the Kramers–Kronig relation (KKR)

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi} \int_0^{\infty} \frac{\Omega}{\Omega^2 - \omega^2} \varepsilon_2(\Omega) d\Omega. \quad (21.16)$$

At low frequency (radio, microwave), in the absence of magnetic fields, $\varepsilon_1 \approx \varepsilon(0)$ and ε_2 , which is responsible for dielectric loss, is small. At optical wavelengths, from far-IR to soft X-rays, the basic quantity is the complex refractive index $\sqrt{\varepsilon} = n + ik$. The real refractive index n , which is responsible for wave propagation properties, and the extinction index k , responsible for the field attenuation, are referred to as *optical constants*. They are related to the dielectric constant via:

$$\begin{aligned} \varepsilon_1 &= n^2 - k^2, \\ \varepsilon_2 &= 2nk, \\ n &= \sqrt{\frac{(\varepsilon_1^2 + \varepsilon_2^2)^{1/2} + \varepsilon_1}{2}}, \\ k &= \sqrt{\frac{(\varepsilon_1^2 + \varepsilon_2^2)^{1/2} - \varepsilon_1}{2}}. \end{aligned} \quad (21.17)$$

21.1.4 Theory of Electrical and Optical Properties

Boltzmann–Equation Approach

The response of carriers in a band to perturbations away from the thermal–equilibrium state, such as applied electric and magnetic fields or impinging electromagnetic radiation, is described by the deviation of the carrier distribution function $f_s(\mathbf{k}, \mathbf{r}, t)$ from the equilibrium Fermi–Dirac distribution $f_0[E_s(\mathbf{k})]$. The current density equals $\mathbf{j}_s = q \int \mathbf{v}_s(\mathbf{k}) f_s(\mathbf{k}, \mathbf{r}, t) d\mathbf{k}$, where $\mathbf{v}_s(\mathbf{k}) = \frac{\partial E_s(\mathbf{k})}{\hbar \partial \mathbf{k}}$ is the microscopic carrier velocity and the integration is performed over the BZ. The process that balances the external perturbations is scattering of carriers by lattice vibrations (phonons), impurities and other carriers. Impurity scattering dominates transport at low temperatures and remains important at room temperature for moderate and high doping levels, although carrier–carrier scattering also becomes appreciable. Under appropriate conditions, one being that $\hbar\omega \ll \bar{E}$ (where \bar{E} is the average carrier kinetic energy), $f_s(\mathbf{k}, \mathbf{r}, t)$ satisfies the quasi-classical Boltzmann kinetic equation. In the opposite, quantum, range, radiation influences the scattering process. Generalized kinetic equations, which interpolate between the quasi-classical and quantum regimes, have also been derived [21.34].

There exist various methods of solving the quasi-classical Boltzmann equation. The relaxation-time method, variational method [21.35–37] for low electric fields, and displaced–Maxwell–distribution approximation [21.38] for high electric fields, were used in early studies. In the last three decades the Monte Carlo technique [21.39], which overcomes limitations inherent to these theories and allows one to calculate subtle details of the carrier distribution, has been applied to various semiconductors, including crystalline Si [21.14]. If $\hbar\omega \ll \bar{E}$, the kinetic and optical characteristics are calculated well using transition probabilities between carrier states, with the radiation quantum absorbed or emitted [21.40]. The most problematic is the interme-

Table 21.4 Parameters of the phonon modes in crystalline Si

Mode	Energy (K)				Sound velocity
LO	760	700–735 ^a	560	580	
TO	760	–	630–690 ^a	680	
LA	0	240–260	500–510 ^a	580	8.99×10^5 cm/s
TA	0	140–160	210–260	220	5.39×10^5 cm/s
q	Γ	Δ	S	X	

^a [21.19]

mediate range – generalized kinetic equations have been shown to recover the extreme ranges, but no working methods have been developed for solutions at $\hbar\omega \approx E$, to the best of our knowledge. The relaxation-time approximation has proved to work well in many cases of scattering in Si. In this framework, the basic quantity is the relaxation-time tensor $\tau_s(E)$, where s is the index indicating the conduction or valence band, and E is the carrier kinetic energy. For electron valleys, $\tau_c(E)$ has the same symmetry as the respective effective-mass tensor – it is diagonal with principal values of, e.g., τ_l , τ_t , τ_t for $\langle 100 \rangle$ etc. For scalar holes $\tau_{1,2}(E)$ can only be introduced in the isotropic-bands approximation. The mobility and the Hall factor for weak electric fields in the relaxation-time approximation are given by

$$\mu_e(\omega) = \frac{q}{3} \left\langle \frac{\tau_l/m_l}{1 - i\omega\tau_l} + \frac{2\tau_t/m_t}{1 - i\omega\tau_t} \right\rangle, \\ r_e = \frac{3 \langle 2\tau_l\tau_t/m_l m_t + (\tau_t/m_t)^2 \rangle}{\langle \tau_l/m_l + 2\tau_t/m_t \rangle^2}; \quad (21.18)$$

$$\mu_h(\omega) = \frac{q}{1 + \beta} \left\langle \frac{\tau_1/m_1}{1 - i\omega\tau_1} + \frac{\beta\tau_2/m_2}{1 - i\omega\tau_2} \right\rangle, \\ r_h = \frac{(1 + \beta) \langle (\tau_1/m_1)^2 + \beta(\tau_2/m_2)^2 \rangle}{\langle \tau_1/m_1 + \beta\tau_2/m_2 \rangle^2}, \quad (21.19)$$

where the angular brackets indicate averaging with the weight $-E^{3/2} f'_0(E)$, $\beta = (m_{d2}/m_{d1})^{3/2}$ is the density ratio of light holes to heavy holes, and the option for a nonparabolic band is retained.

Lattice Scattering

Deformational phonons – longitudinal, transverse acoustical (LA, TA) and optical (LO, TO) – mediate carrier–lattice scattering in Si. The phonon modes are presented in Table 21.4, where \mathbf{q} is a point in the phonon BZ. The points Δ and S correspond to the scattering process, where the electron transits between the bottoms of two perpendicularly (f) and parallel (g) oriented valleys, respectively. The phonon energies are precise at the points Γ and X, as determined by neutron-scattering techniques [21.42], but uncertain at Δ and S, since in this case only estimation and fitting methods were available.

The rigid- and deformable-ion lattice models have been used to obtain the carrier–phonon interaction for

electrons [21.43] and holes [21.44]. A deformation-potential theory of the interaction with long-wavelength phonons, which takes the crystal symmetry and band structure fully into account in a phenomenological manner, has been developed. This theory deduces two constants, Ξ_u , Ξ_d , for the conduction band [21.45] and four, a , b , d , d_{opt} , for the valence band [21.46,47], which are presented in Table 21.5. The deformation-potential theory has been used to calculate the acoustic scattering-limited mobility in n-Si [21.48] and p-Si [21.49]. Later, optical-phonon scattering, along with an approximate valence-band spectrum instead of (21.5), were taken into account [21.13, 14].

The matrix elements of electron–phonon interaction between wave functions of different valleys are not taken into account by the deformation-potential theory. For inter-valley transitions, other than the three marked in Table 21.4 by the superscript ‘a’, the matrix elements calculated at the valley-bottom wave vectors are zero [21.43, 50]. The actual scattering probabilities are never zero; for those forbidden by selection rules [21.43, 50] one should take into account the wave-vector offset at the final scattering state, which gives nominally small, but unknown values. Several inter-valley scattering models have been tried to fit the theoretical formulas to the mobility data in lightly doped n-Si: with one allowed TO and one forbidden TA phonon [21.51], one allowed TO phonon [21.52] and more involved combinations of the transitions [21.53, 54]. The scattering of electrons by long-wavelength optical phonons, regarded as a cause of drift-velocity saturation at high electric fields [21.38], is forbidden in Si [21.43]. In n-Si, by all accounts, the cause may be the allowed g-phonon (Table 21.4) scattering [21.54].

Impurity Scattering

There are two types of impurity scattering – by ionized and neutral impurities. The latter is the dominant impurity scattering for uncompensated, light or moderate shallow-impurity doping, at low T . In samples doped with deep impurities, neutral-impurity scattering may also show up. At elevated T , when shallow impurities are increasingly ionized, ionized-impurity scattering is the dominant impurity scattering and may compete with

Table 21.5 Deformation-potential parameters [21.41]

$T(\text{K})$	$\Xi_u(\text{eV})$	$\Xi_d + 1/3\Xi_u - a(\text{eV})$	$ b (\text{eV})$	$ d (\text{eV})$
80	8.6 ± 0.2	3.8 ± 0.5	2.4 ± 0.2	5.3 ± 0.4
295	9.2 ± 0.3	3.1 ± 0.5	2.2 ± 0.3	–

phonon scattering, depending on $N_{d(a)}$. Lastly, in heavily doped samples, where the impurities are ionized for all T , ionized-impurity scattering is dominant up to 300 K.

Early theories of impurity scattering were developed for carriers in parabolic bands, scattered by hydrogen-like centers. For neutral impurities the s -scattering cross section [21.55] and a cross section that takes allowance of the scattered carrier's bound state [21.56] were adopted. For ionized centers, use was made of the Coulomb scattering cross section, cut off at a small angle depending on $N_{d(a)}$ [21.57], and the screened Coulomb potential cross section, calculated in the Born approximation [21.58–60]. These theories consider scattering by the donors and acceptors on an equal footing. The use of the Conwell–Weisskopf formula for $\tau_s(\varepsilon)$ declined towards the end of the 1950s, while the corresponding Brooks–Herring (BH) formula became widespread, mostly due the consistency of its derivation, even though none of the assumptions for its validity are completely satisfied. This formula was corrected [21.61], on account of the band carrier's degeneracy, compensation and screening by carriers on impurity centers. The discrepancy between experiment and the BH formula, revealed during three decades of studies, have been thoroughly analyzed [21.62].

Modifications and developments made to overcome the drawbacks of the BH formula are worth mentioning. Taking the multi-valley band structure into account did not invalidate the relaxation-time method as such, but resulted in essentially different $\tau_1(E)$ and $\tau_1(E)$ [21.63, 64]. These formulas have also been discussed [21.65] in light of the scattering anisotropies measured in n-Si [21.66]. To overcome limitations of the Born approximation exact, although limited only to the standard band, phase-shift analysis was employed [21.67]. Including a non-Coulomb part of the impurity potential [21.68] made it possible to explain in part the difference in mobility of n-Si samples doped with different donors [21.25]. Lastly, Monte Carlo simulations of the impurity scattering, improving the agreement between theory and experiment for n-Si, have recently been reported [21.69]. The ionized-impurity scattering in p-Si was considered in the approximation of isotropic hole bands [21.70]. We are not aware of any theoretical papers on the subject, which used anisotropic energy spectra given by (21.26) in the case of p-Si.

Carrier–Carrier Scattering

Carrier–carrier scattering becomes important as n or p increases, along with increasing N_d^+ or N_a^- . The relaxation-time concept does not apply for

this mechanism. Carrier–carrier collisions redistribute the carrier's energy in a chaotic manner that was presumed to cause a decrease in the net mobility due to other mechanisms [21.71]. For the standard band, the effects of electron–electron scattering were modeled using the variational method, which predicted a $\approx 30\%$ reduction in the ionized-impurity scattering-limited mobility [21.72]; close results were obtained using another, quite different, method [21.73]. Hole–hole scattering and electron–hole scattering were also considered [21.74] in the standard band. Due to ignorance of the specific band-structure features, the results of these papers had limited relevance to Si. The effect of electron–electron scattering has been recast [21.75] for the multi-valley band structure using the generalized Drude approximation (GDA). At DC, the GDA corresponds to the zeroth-order approximation of the variational method, which highly overestimates [21.72] the effect considered.

Dielectric Constant

In Si the current carriers are well decoupled from the host electrons, so the Maxwell equations result in a unique decomposition of the dielectric constant

$$\varepsilon(\omega) = \varepsilon_L(\omega) + \varepsilon_C(\omega), \quad \varepsilon_C(\omega) = i \frac{4\pi\sigma(\omega)}{\omega}. \quad (21.20)$$

Here $\varepsilon_L(\omega)$ is the host contribution, which is indirectly influenced by the carriers. The direct effect of the carriers on the dielectric constant is the conductivity contribution $\varepsilon_C(\omega)$. As seen from (21.20), doped Si behaves at DC as a metal. The asymptote at high frequencies (IR for Si) is $\varepsilon_{C,s}(\omega) \approx -(\Omega_{pl,s}/\omega)^2$, where $s = e$ or h , and $\Omega_{pl,s}$ are the bare plasma frequencies given by

$$\begin{aligned} \Omega_{pl,e}^2 &= \frac{4\pi q^2 n}{m_{ce}}, \quad \frac{1}{m_{ce}} = \frac{1}{3} \left\langle \frac{1}{m_1} + \frac{2}{m_t} \right\rangle, \\ \Omega_{pl,h}^2 &= \frac{4\pi q^2 p}{m_{ch}}, \quad \frac{1}{m_{ch}} = \frac{1}{1+\beta} \left\langle \frac{1}{m_1} + \frac{\beta}{m_2} \right\rangle, \end{aligned} \quad (21.21)$$

resulting from (21.18, 19, 20), irrespective of the scattering model. Due to this asymptote and (21.20) $\varepsilon_1(\omega)$ should become zero at some frequency $\omega_{pl,s}$ and doped Si should behave optically as a dielectric at $\omega < \omega_{pl,s}$ and as a metal at $\omega > \omega_{pl,s}$. True plasma frequencies are estimated roughly as $\omega_{pl,s} \approx \Omega_{pl,s}/n_L$, where $n_L = \text{valrange}3.423.44$ is the Si host refractive index in the IR. For the parabolic bands, m_{ce} has already been presented in Table 21.2, and $m_{ch} = (0.33\text{--}0.39)m_0$ us-

ing the data of Table 21.2, with the assumption that $m_{1,2} = m_{c1,2}$.

General formulas for $\varepsilon_C(\omega)$ are rather involved because of the averaging over E they contain, and so are rarely used. The Drude formula

$$\varepsilon_{C,s}(\omega)|_{\text{Drude}} = \frac{i}{\omega} \cdot \frac{\Omega_{\text{pl},s}^2}{\gamma_s - i\omega}, \quad (21.22)$$

where γ_s are adjustable phenomenological damping parameters, is often employed instead [21.76]. To match the behavior of $\varepsilon_C(\omega)$ at $\omega \rightarrow 0$ one should put $\gamma_s = 1/\tau_{0,s}$ in (21.22), where

$$\begin{aligned} \tau_{0,e} &= \frac{m_{ce}}{3} \left\langle \frac{\tau_1}{m_1} + \frac{2\tau_1}{m_1} \right\rangle; \\ \tau_{0,h} &= \frac{m_{ch}}{1+\beta} \left\langle \frac{\tau_1}{m_1} + \frac{\beta\tau_2}{m_2} \right\rangle \end{aligned} \quad (21.23)$$

are the DC mobility relaxation times. Such an adjustment was shown to work poorly in n-Si [21.77]. On the other hand, putting $\gamma_s = \gamma_{\infty,s}$ in (21.22), where

$$\begin{aligned} \gamma_{\infty,e} &= \frac{m_{ce}}{3} \left\langle \frac{1}{m_1\tau_1} + \frac{2}{m_1\tau_1} \right\rangle; \\ \gamma_{\infty,h} &= \frac{m_{ch}}{1+\beta} \left\langle \frac{1}{m_1\tau_1} + \frac{\beta}{m_2\tau_2} \right\rangle, \end{aligned} \quad (21.24)$$

allows one using the Drude formula to match *two* leading at $\omega \rightarrow \infty$ terms in the power series expansion of $\varepsilon_C(\omega)$ with respect to ω^{-1} . Thus (21.22) with the above adjustments may serve as an overall interpolation if the *high-frequency* relaxation time, $\tau_{\infty,s} = 1/\gamma_{\infty,s}$ turns out to be close to $\tau_{0,s}$. A Drude formula, empirically adjusted in IR has been devised for n-Si [21.78].

Using $\tau_{\infty,s}$ instead of $\tau_{0,s}$ in the mobility is a prerequisite for GDA at DC. As discussed above, the Boltzmann-equation-based formulas are valid in the range $\lambda \gg \lambda_q = hc/\bar{E}$. For nondegenerate carriers $\lambda_q(\text{cm}) \approx 1.4388/T$, while in Si for n or p up to $\approx 10^{20} \text{ cm}^{-3}$ the carrier degeneracy (if present) stops much below 300 K, so the validity of the $\varepsilon_C(\omega)$ formulas considered is restricted to the far-IR and longer wavelengths. To properly describe $\varepsilon_C(\omega)$ in the near- and mid-IR range, ω -dependent GDA has been suggested [21.79]. In this approximation one replaces γ_s in (21.22) by an ω -dependent damping $\gamma_s(\omega)$, which is then determined by comparison of the first imaginary term in the expansion of the thus-generalized Drude formula, i.e. $(\Omega_{\text{pl},s}/\omega)^2 \gamma_s(\omega)/\omega$, with $\varepsilon_2(\omega)$ calculated using the methods of transition probabilities or perturbations for correlation functions.

In Si, unlike semiconductors with ionic bonds (e.g. $\text{A}^{\text{III}}\text{B}^{\text{V}}$), the elementary cell has no dipole moment and hence no quasi-classical optical-phonon contribution in $\varepsilon_L(\omega)$ is present. However, several weak IR absorption bands, attributed to two-phonon interaction of light with the Si lattice, are observed. Of these bands the most prominent is that peaked at $16.39 \mu\text{m}$, which undergoes about a twofold increase in absorption upon increasing the temperature from 77 to 290 K. Comparable to that, the $9.03 - \mu\text{m}$ absorption band, observed in pulled Si crystals, was attributed to Si–O bond stretching vibrations [21.80]. At $\lambda < 1.2 \mu\text{m}$, where $\varepsilon_L(\omega)$ dominates the dielectric constant irrespective of the doping, $\varepsilon_{2L}(\omega)$ is accounted for by inter-band electronic transitions. The spectral bands of $\varepsilon_{2L}(\omega)$ correspond to the absorption of photons with energies close to the band gaps. Bare indirect-band-gap transitions, that necessitates lowest energy, is forbidden, but perturbation correction in the electron–phonon interaction to $\varepsilon_{2L}(\omega)$ suffices to describe the observed indirect absorption band. Due to the phonons the lowest fundamental absorption, although smaller than in direct bands, increases with increasing T . In principle, $\varepsilon_{2L}(\omega)$ may be calculated by the band-structure simulation route (e.g. Kleinman and Phillips [21.6, 8, 21]), especially with the present state-of-the-art theory – $\varepsilon_{1L}(\omega)$ is then calculated using the KKR (21.16). However, applications need fast modeling, and such formulas have been developed for Si [21.81–83]. There are a few distinct doping effects on $\varepsilon_L(\omega)$:

1. Effects on the lowest band edges, both direct and indirect: the Burstein–Moss shift with increasing $n(p)$ in heavily doped samples at the degeneracy due to filling of the conduction (valence) band below (above) E_F ; band-gap shrinkage due to carrier–carrier and carrier–impurity interactions [21.84, 85] that work against the Burstein–Moss shift; and the formation of band tails because of the random potential of impurities (e.g. [21.31]).
2. Effects on higher edges, such as: E_1 (3.4 eV) – due to transitions between the highest valence band and the lowest conduction band along the Λ line in a region from $\pi/4a_0(1, 1, 1)$ to the L point on the BZ edge; and E_2 (4.25 eV) – due to transitions between the valence band at the X point and the conduction band at $2\pi/a_0(0.9, 0.1, 0.1)$ [21.8]. In this case [21.86], the electron–electron interaction plays a small role because carriers are located in a small region of the BZ, different from that where the transitions take place,

and the effect of the electron–impurity interaction is calculated using standard perturbation theory.

3. Absorption due to direct inter-conduction-band (inter-valence-band) transitions specific to the type of doping. In n-Si this is a transition from the lowest conduction band to the band that lies higher at the Δ point but crosses the former at the X point, and which gives rise to a broad absorption band peaked around 0.54 eV [21.87] and tailing of the indirect gap at heavy doping [21.88]; the theory of this contribution to $\varepsilon_{2L}(\omega)$ has been developed [21.85]. In p-Si these transitions are those between the three highest valence bands [21.7]. Absorption due to the $1 \rightarrow 2$ tran-

sition has no energy threshold, and so resembles the usual free-carrier absorption; that due to the $1 \rightarrow 3$ and $2 \rightarrow 3$ transitions appears at $\hbar\omega = \Delta_{so}$. A high-energy threshold, above which the inter-valence-band absorption becomes negligible, exists due to the near congruency of all three valence sub-bands [21.7] at large well k (this is not accounted for by (21.5, 6), which are valid at small k). In contrast, with p-Ge, the manifestation of the inter-valence-band transitions in the reflection, was proposed for p-Si [21.89], but fully reconciled later [21.85, 90]. Kane's theory of the inter-valence-band absorption [21.7] has also been revisited [21.90].

21.2 Electrical Properties

An extensive investigation of basic electrical properties was started 55 years ago, when polycrystalline Si containing B and P, was reported [21.91]. This seminal work was necessarily limited because neither single crystals nor the means for measuring below 77 K were then available. The first papers on single crystals were published five years afterwards. Since then, as techniques for fabricating quality single-crystalline silicon, such as the pulling, e.g. Czochralski (CZ), Teal–Little (TL), and floating-zone (FZ) techniques, became highly developed, many experiments on electrical properties have been published. A number of papers are considered below in historical retrospect. In the accompanying graphs additional, less cited, papers are referenced.

Though the physical mechanisms behind the electrical properties of crystalline Si have been studied and partially understood for a long time, the resulting formulas and procedures are too complicated and time-consuming to be used in electronics device modeling. In this connection, several useful, simplified but accurate, procedures for modeling mobility versus temperature, doping, injection level and electric field strength have been developed. For this issue we refer to points 2–4 in Sect. 21.2.3.

21.2.1 Ohm's Law Regime

Drift Measurements

1. The minority-carrier mobility as a function of N_d , N_a , T and ρ in n- and p-type samples is in the range 0.3–30 Ωcm [21.92]. This cited paper revealed for the first time the inapplicability, at least for holes, of

the simple $T^{-1.5}$ lattice mobility law, and presented curves of ρ versus exhaustion concentration $N = |N_d - N_a|$ in the range $10^{14} \text{ cm}^{-3} \leq N \leq 10^{17} \text{ cm}^{-3}$.

2. Measurements of μ_e in p-type and μ_h in n-type Si on 11 single crystals ranging in ρ from 19 to 180 Ωcm [21.93]. In the purest crystals, in the range 160–400 K, μ_e and μ_h obeyed the dependencies $T^{-2.5 \pm 0.1}$ and $T^{-2.7 \pm 0.1}$, respectively. The conductivity of some of these crystals, measured from 78 to 400 K, provided independent evidence for the temperature dependencies of the mobility quoted above.
3. Room-temperature drift and conductivity plus Hall-effect measurements [21.94] of μ_e and μ_h versus resistivity on an unprecedentedly large number of samples cut from CZ crystals. The largest ρ was above 200 Ωcm . The values of μ_e and μ_h obtained from both experiments in the purest crystals were reported and compared with those obtained by other authors.

Resistivity

and Galvanomagnetic Measurements

1. Room-temperature $\mu_{e(h)}^H$ as a function of $\rho_{e(h)}$ [21.95]. The crystals used were grown from Dupont hyper-pure material, with ρ of 0.01–94 Ωcm for n-type and 0.025–110 Ωcm for p-type samples. Curves of $\mu_{e(h)}^H$ versus $\rho_{e(h)}$ were calculated using the BH and combined-mobility [21.71] formulas for $m_{e(h)} = m_0$ and compared with experimental curves.
2. The first extensive experimental study of electrical conductivity and the Hall effect in TL silicon [21.96]. The properties were measured at

- temperatures of 10–1100 K on six arsenic-doped n-type samples, and one undoped, plus five boron-doped, p-type samples, covering the range from light ($N = 1.75 \times 10^{14}$ and $3.1 \times 10^{14} \text{ cm}^{-3}$) to heavy ($N = 2.7 \times 10^{19}$ and $1.5 \times 10^{19} \text{ cm}^{-3}$) doping. Compensation by unknown acceptors (donors) occurred in four lightly and moderately doped n(p)-type samples. A deviation of the lattice mobility from the $T^{-1.5}$ dependence was reported for both electrons and holes. Curves of $\mu_{e(h)}^H$ against $\rho_{e(h)}$ at 300 K were computed in the same way as by Debye and Kohane, but incomplete ionization of impurity centers was additionally taken into account.
3. First systematic study of μ_e and μ_h versus N_d and N_a , respectively, at $T = 300 \text{ K}$ [21.97]. Measurements were taken with several group V and group III impurities up to 6×10^{19} and $6 \times 10^{18} \text{ cm}^{-3}$ for n- and p-Si, respectively. Impurity concentrations were obtained by radioactive tracers or from thermal neutron activation analysis; μ_e and μ_h were calculated from these data by considering the N_d^+ and N_a^- percentages. The combination with measured $\mu_{e(h)}^H$ resulted in $r_{e(h)}$ values in agreement with theory. A comparison with a BH-formula-based theory yielded semiquantitative agreement for μ_e^H , while measured values of μ_h^H proved to be much smaller than the theoretical values.
 4. Galvanomagnetic effects in p-Si: ρ and R^H versus T and B [21.98] and MR [21.99]. Boron-doped samples cut from CZ crystals were used. In the first paper four samples, two with $\rho(300 \text{ K}) = 35 \Omega\text{cm}$ and two with $\rho(300 \text{ K}) = 85 \Omega\text{cm}$, were measured in the range 77–320 K. The dependence $\mu_h \propto T^{-2.7 \pm 0.1}$ at $B = 0$, as observed by Ludwig and Watters, was typical of the results obtained on all the samples; r_h was observed to exhibit a weak linear decrease with T in the range 200–320 K, and to be almost entirely independent of B up to $B = 1.3 \text{ T}$ in the temperature interval studied. The dependence of MR on the relative directions of current, fields and crystallographic axes was studied at 77 K and 300 K as a function of B . Large values of longitudinal MR, as large as the transverse effects in some cases, were observed, contradicting the only calculations available at that time [21.9]. To obtain data sufficient for constructing a more satisfactory model, the above study was continued in the second paper on 10 more samples, with ρ of 0.15–115 Ωcm . Measurements of three MR coefficients were carried out at a number of temperatures in the range 77–350 K. The results showed a marked dependence of the band structure and scattering anisotropies on the temperature, yet no definite model of these effects was arrived at.
 5. A comparative study of mobility in pulled and FZ crystals [21.100]. The question of the dependence of the intrinsic mobility on temperature was recast. The authors found that in FZ, contrary to pulled crystals, μ_e^H followed the $T^{-1.5}$ law in the range 20–100 K, although μ_h^H still displayed a different, viz. T^{-2} , variation with temperature. It was argued that such a disagreement with the work of Morin and Maita was due to the large content, up to 10^{18} cm^{-3} , of oxygen impurities in the pulled crystals they used, which resulted in scattering that obscured the phonon scattering.
 6. Solid analysis of electrical properties of n-Si with respect to: the ionized-impurity scattering in isotropic approximation [21.101], scattering anisotropies [21.66] and lattice scattering [21.51]. The measurements were made from 30 to 100–350 K using a set of P-doped, B-compensated, n-type samples of rather wide impurity content, yet in the range from light to moderate doping ($N_d = 4.5 \times 10^{15} \text{ cm}^{-3}$ at most). The purest samples were cut from FZ crystals while others were from CZ crystals. These authors developed a sophisticated, but robust, method of determining N_d and N_a by analysis of the R^H versus T data. With this method, in the first paper they obtained curves of μ_e versus T in the range 30–100 K, which were used to test the BH formula. In comparing the formula with the data, correction to the observed μ_e because of the phonon-scattering contribution was necessary. The BH formula was shown to provide a good quantitative description of the data when $m_e = 0.3m_0$ was used, provided that ion scattering was not too strong. When ion scattering was dominant, viz. in moderately doped samples at low T , they observed a discrepancy between the theory and data, which was attributed to electron–electron interaction. For the purpose of detecting scattering anisotropy, the MR coefficients were measured in the second paper on several relatively pure ($N_d = 8.0 \times 10^{14} \text{ cm}^{-3}$ at most) samples. The results indicated that $\tau_l/\tau_t \approx 0.67$ and $\tau_l/\tau_t > 1$ for acoustic-phonon and ionized-impurity scattering, respectively. The inter-valley phonon scattering, important at higher T , proved to be isotropic. In the third paper lattice scattering was treated. A model assuming inter-valley scattering by two-phonon modes, in addition to the intra-valley acoustic-phonon scattering, was applied to the re-

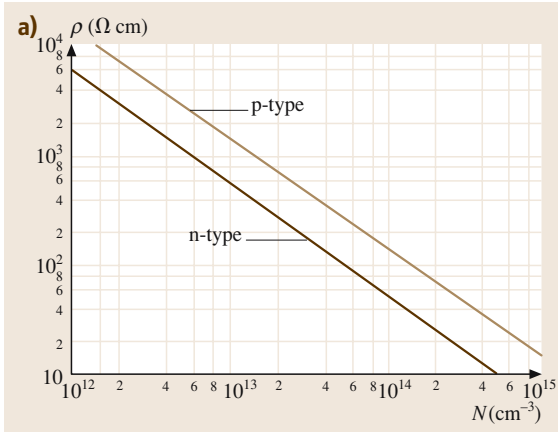


Fig. 21.5 (a) Room-temperature resistivity of Si versus N (defined in the text) for overlapping ranges of doping: (a) light to moderate;

sults of the electrical and MR measurements in the purest samples. By using the coupling constants with the phonon modes as fitting parameters, fairly good agreement was achieved between the model and experimental data on curves of μ_e versus T in the range 30–350 K, though one of the inter-valley phonon energies corresponded to a forbidden transition (Table 21.4).

7. Experimental verification of the anisotropic ion-scattering theory [21.102]. Measurements of saturated longitudinal MR and μ_e were made in fields up to 9 T at 78 K on a series of [111]-oriented phosphorus-doped samples with N_d in the range 2×10^{13} – 6×10^{16} cm^{-3} . A quantitative analysis was made that involved combined re-

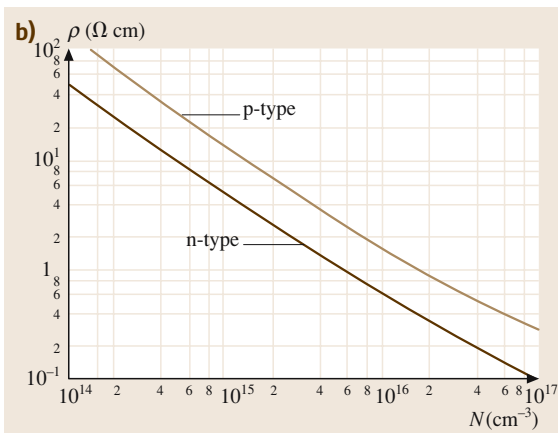


Fig. 21.5 (b) moderate;

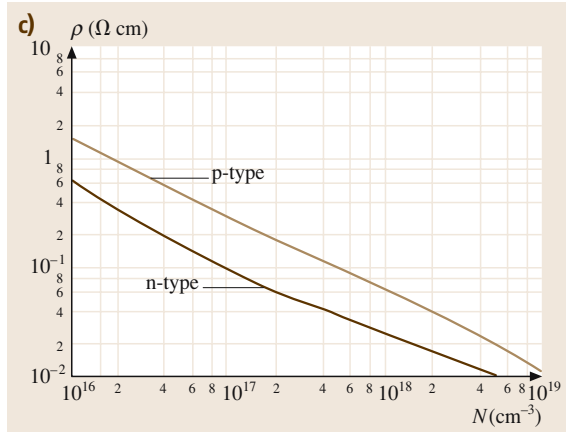


Fig. 21.5 (c) moderate to heavy;

laxation times from impurity, intra-valley, and inter-valley lattice scattering. The data agreed very well with the theory [21.103]. In samples with $N_d > 10^{16}$ cm^{-3} the neutral-impurity scattering effect was observed.

8. Refined analysis of the Hall effect and mobility in n-Si [21.53]. μ_e and n , as determined from the high-field Hall effect, were numerically analyzed for a series of n-type samples doped with Sb, P, and As. The calculations were based on the general treatment for an anisotropic parabolic band [21.48]. Lattice-scattering parameters for the inter-valley and acoustic modes were determined from a comparison of the results between theory and experiment, using as many as four inter-valley phonons. The conclusions supported the two-phonon Long model. Ionized-impurity scatter-

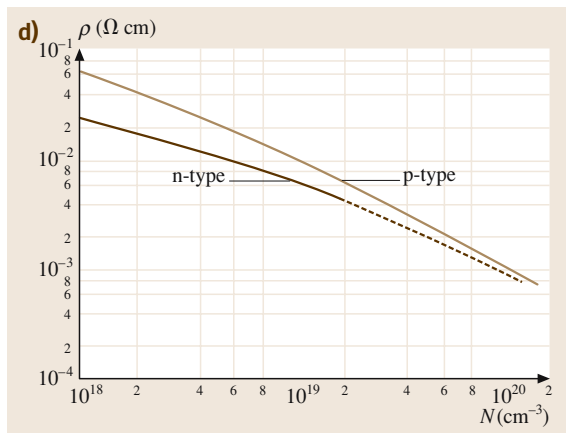


Fig. 21.5 (d) heavy. After [21.24] with permission

ing was calculated from the BH formula and a newer theory [21.103] and compared with experiment in favor of the latter. Neutral-impurity scattering proved

to be temperature-dependent, unlike the available model [21.55].

9. The Hall factor in n-Si versus $N_I = N_d^+ + N_a^-$ [21.106]. The experimental value of r_e was obtained from the ratio of $R_e^H(0)$ to $R_e^H(\infty)$, determined by independent methods at 77 and 350 K over a wide range of N_I . Samples were cut from non-oxygenated ingots grown in vacuum, in which $R_e^H(0)$ was measured by the standard method (Sect. 21.2). Discs were then cut from the same samples, in proximity to Hall probes, and their total resistances (given by $BR_e^H(\infty)/d$ at $\mu_e B/c \gg 1$ (Sect. 21.2), where d is the disc width) were measured in fields up to $B = 3.5$ T. To check these results, standard measurements of R_e^H in pulsed fields (Sect. 21.2) up to $B = 15$ T were made. The values of r_e were calculated from theory [21.103]. Good agreement of experiment with theoretical calculations was obtained.

Minority Carriers – Miscellaneous

Minority carriers play a crucial role in silicon-based electronic devices such as bipolar junction transistors (BJT) and solar cells, so their charge-transport parameters – concentration, recombination time, mobility and diffusion coefficient – are important for device modeling. We do not present a literature overview nor graphical and tabular material here, as we do for the case of the majority carriers, and refer the reader directly to the review papers outlined in points 5 and 6 of Sect. 21.2.3.

Irvin Curves

The curves of $\rho(300\text{ K})$ versus N , over the entire range up to the solubility limit, compiled using published data [21.104, 107] are shown in Fig. 21.5a–d, with the high-doping range shown Fig. 21.6a–d. The resistivity decreases with increasing N since the conductivity is proportional to $n(p)$ while $n(p) = N$ at $T = 300\text{ K}$. However, at large N the decrease is slower than $\propto N^{-1}$ due to the notable dependence of $\mu_{e(h)}$ on doping (see below).

Mobility and Hall Effect Versus Resistivity and Doping

Many studies have reported μ versus ρ graphically. Figure 21.7a–d summarize this information obtained before 1965. For the Hall measurements the drift mobility was recalculated from the Hall value using the Hall factor for the standard band (Sect. 21.1.3). The dependencies of $\mu_{e(h)}$ on N , $N_{d(a)}$ and on N_I at low T are more physical and allow comparison with available theories. Exam-

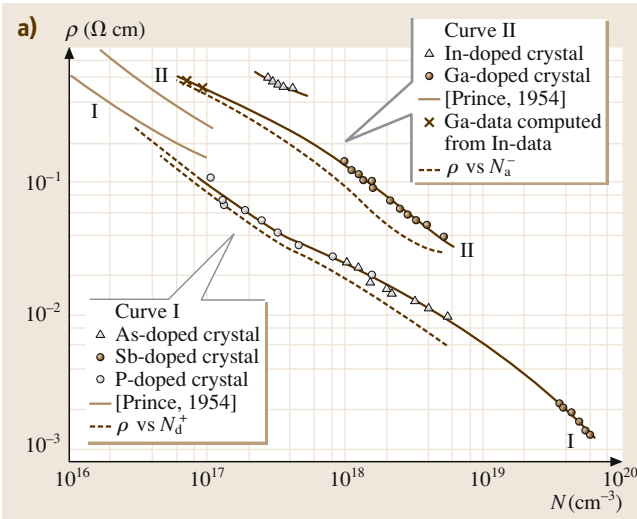


Fig. 21.6 Room-temperature resistivity of highly doped Si versus: (a) N in compensated n- and p-type crystals; the curves corrected for ionized impurities content are also presented [21.97] with permission;

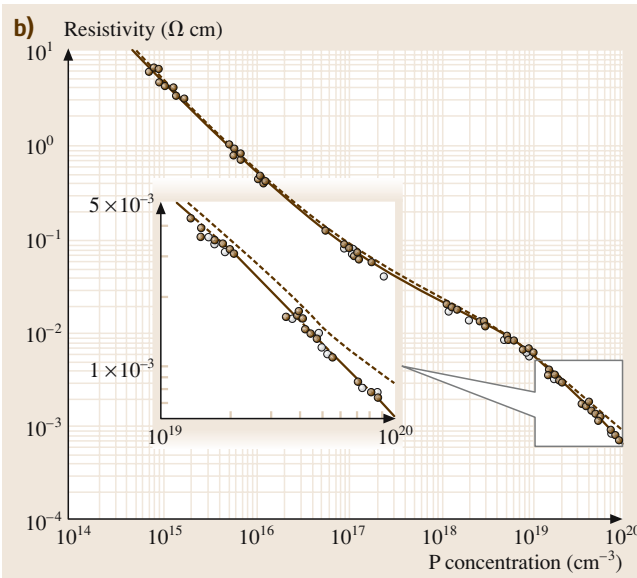


Fig. 21.6 (b) P content in uncompensated P-doped Si. Dashed curve is due to Irvin [21.104]. P concentration by: solid circle – neutron activation analysis, open circle – Hall-effect measurement. Inset – heavily doped range enlarged. After [21.105] with permission

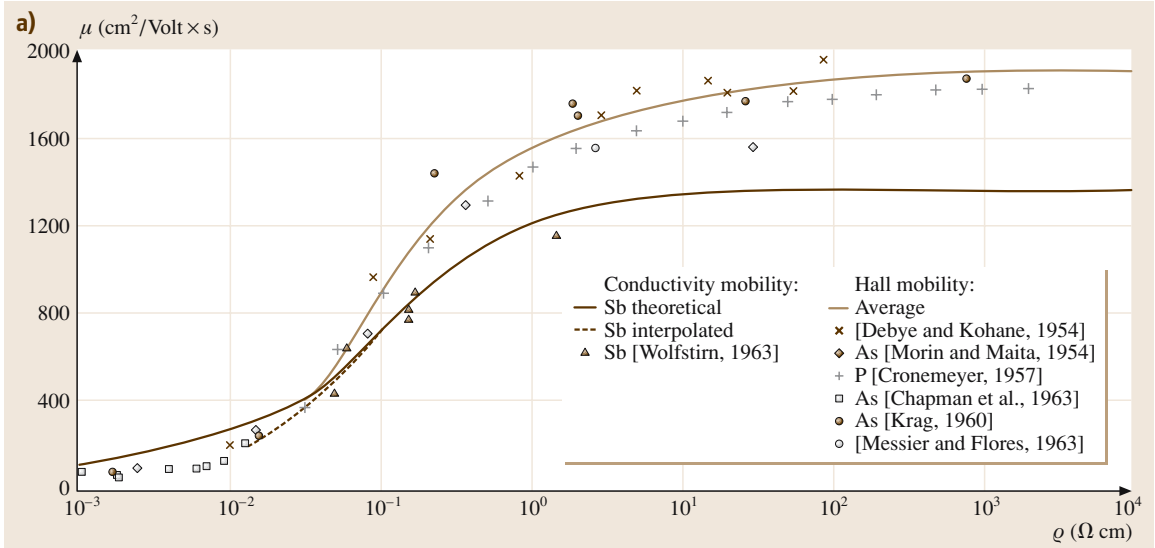


Fig. 21.7 Mobility versus resistivity in Si at room temperature compiled from different authors: **(a)** Electron mobility. After [21.24] with permission;

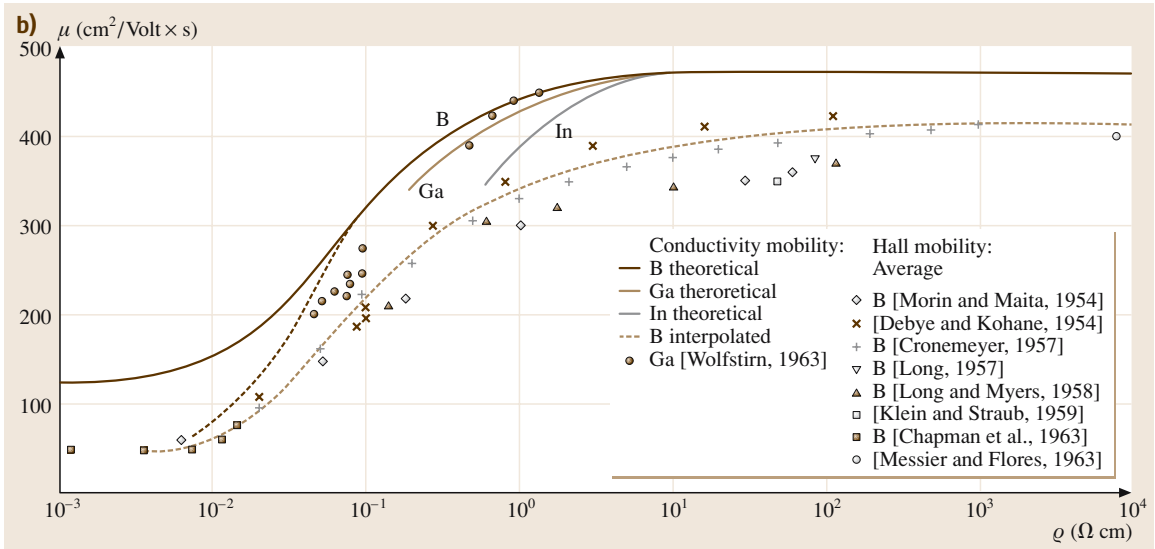


Fig. 21.7 (b) Hole mobility. After [21.24] with permission;

ples of such dependencies obtained by different authors are presented in Fig. 21.8a–e. Several phenomenological expressions for $\mu_{e(h)}(N, 300\text{ K})$ have been given in the literature (Sect. 21.2.3) in agreement with experimental data on charge transport in Si. We consider here two of them, for obvious reasons of space. The first [21.108]

reads

$$\mu(N_i) = \mu_{\min} + \frac{\mu_{\max} - \mu_{\min}}{1 + (N_i/N_{\text{ref}})^\alpha}, \quad (21.25)$$

where $i = a, d$, and the other fitting parameters are given in Table 21.6, the electron column of which corre-

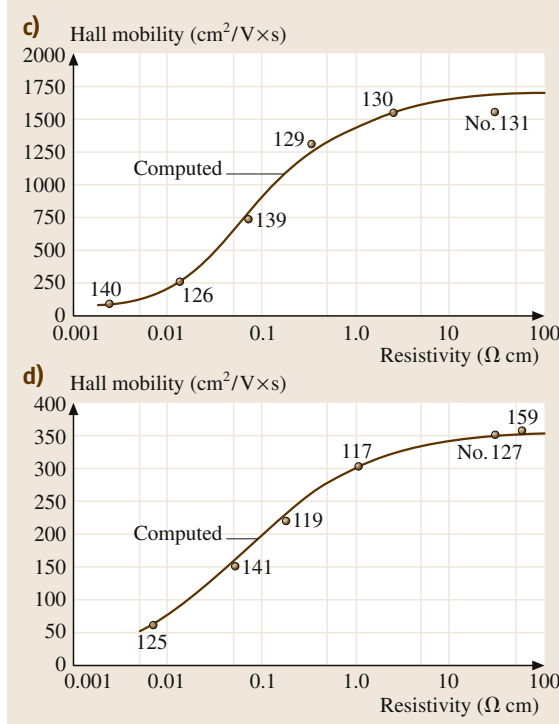


Fig. 21.7 (c) Hall mobility of electrons in Si:As. After [21.96] with permission; (d) Hall mobility of holes in Si:B. After [21.96] with permission. Numbers on graphs (c) and (d) denote samples, details of which are given in Table 21.10

sponds to a later best fit for Si:P [21.109], while the hole column is due to *Caughey and Thomas* [21.108]. The particular case of (21.25) with $\mu_{\min} = 0$, $\alpha = 0.5$, $N_{\text{ref}} = 10^{17} \text{ cm}^{-3}$ [21.110] has also proved useful. The second expression for Ohm's mobility versus impurity concentration,

$$\mu(N_i) = \frac{\mu_0}{\sqrt{1 + \frac{(N_i/N_{\text{eff}})S}{N_i/N_{\text{eff}} + S}}}, \quad (21.26)$$

where N_{eff} and S are fitting parameters, follows from the weak electric field limit of the more general phenomenological expression (21.31) for v_d versus E and N_i . The best-fit parameters for (21.26) can be found in Table 21.7. Fits to (21.25) and (21.26) are used in Fig. 21.8a–c. Physically, μ_{\max} in (21.25) and μ_0 in (21.26) are $\mu_{e,hL}(300 \text{ K})$, the intrinsic mobility (see below) at room temperature. Experimental data on this parameter due to different authors are presented in Table 21.8. Figure 21.8d shows a com-

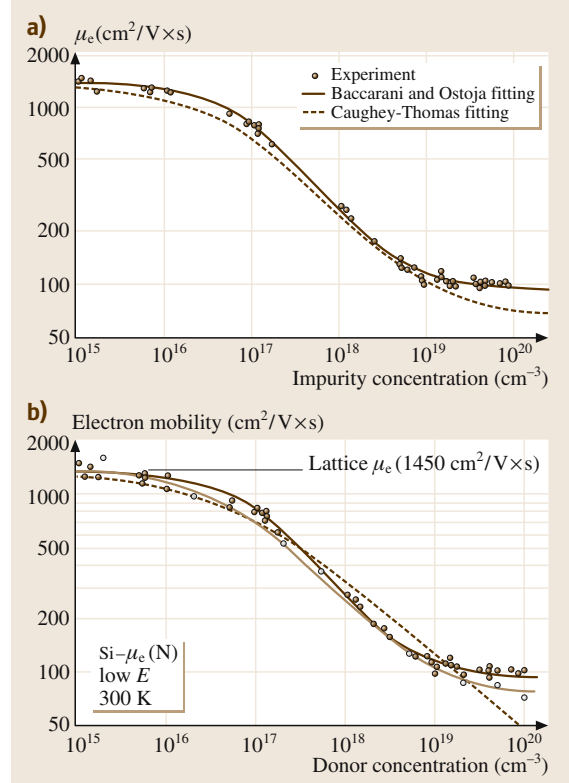


Fig. 21.8 Mobility versus doping in Si: (a) Room-temperature electron mobility in n-type Si: P versus P concentration, *solid square* – experimental data [21.105], *full line* – the best fit due to Baccarani and Ostojia ((21.25), Table 21.6) and *dashed line* – fitting of Irvin's data by (21.25) due to Caughey and Thomas [21.109], with permission; (b) Summary of data on and fits to room-temperature electron mobility versus donor concentration in n-Si: *open circle* – compiled experimental data [21.104], *solid circle* – experimental data [21.105], *full line* – the best fit due to Baccarani and Ostojia ((21.25), Table 21.6), *broken line* – the best fit by (21.25) due to Hilsum and Gummel and *dash-dotted line* – the best fit due to Scharfetter and Gummel ((21.26), Table 21.9) [21.14], with permission;

parison between the mobility of holes, as minority and majority carriers, plotted against the concentration of majority impurities. At the same concentration the minority-hole mobility is strikingly larger than the majority-hole mobility. This effect may be due to the difference between scattering by attractive and repulsive centers as well as to hole–hole interaction. The variation of $\mu(N, 300 \text{ K})$ with shallow donor and accep-

Fig. 21.8 (c) Summary of data on and fits to room-temperature hole mobility versus acceptor concentration in p-Si, *open circle* – compiled experimental data [21.104], *solid circle* – lattice mobility, *full line* – fitting of Irvin's data by (21.25) due to Caughey and Thomas, and *dash-dotted line* – the best fit due to Scharfetter and Gummel ((21.26), Table 21.9) [21.14], with permission; **(d)** Mobility of holes as being minority carrier (*symbols and dashed line*), measured on devices by different authors indicated in captions, and majority carrier (*full line*) due to Thurber et al. After [21.111] with permission, the references are found in this review;

tor species, noted in Sect. 21.1.4, is shown in Table 21.9. As can be seen, the effect – the larger $E_{d(a)}$ the lower $\mu_{e(h)}$ – is not drastic but quite noticeable. At low temperatures the impurity-scattering-limited mobility depends on the density of ionized centers, as predicted by the BH formula (or its modifications), e.g. for electrons

$$\mu_{eI} = \frac{\eta T^{3/2}}{N_I (\ln b - 1)}, \quad (21.27)$$

$$\eta = \frac{2^{7/2} k_B^{3/2} \varepsilon'^2}{\pi^{3/2} q^3 m^{*1/2}},$$

where m^* is an adjusted effective mass, $\varepsilon' = \varepsilon(0)$ is the static dielectric constant and

$$b = \frac{24\pi m \varepsilon' (k_B T)^2}{n' q^2 h^2}, \quad (21.28)$$

$$n' = n + (n + N_a) \left(1 - \frac{n + N_a}{N_d}\right).$$

For compensated n-Si $n \ll N_a$ at sufficiently low T , so $N_I = 2N_a$ and $n' = N_a(N/N_d) = N_I(1 - K)/2$, where K is the compensation ratio. Thus, disregarding the weak logarithmic dependence, μ_{eI}^{-1} proves to be linear with N_I . Such dependence is seen in the experimental data shown in Fig. 21.8e.

The Hall coefficient for lightly compensated n-Si as a function of B is shown in Fig. 21.9a. The curve for

Table 21.6 Best fit parameters for the Caughey–Thomas formula (21.25)

Parameters	Electrons	Holes	Units
μ_{\min}	92.0	47.7	$\text{cm}^2\text{V}^{-1}\text{s}^{-1}$
μ_{\max}	1360	495	$\text{cm}^2\text{V}^{-1}\text{s}^{-1}$
N_{ref}	1.3×10^{17}	6.3×10^{16}	cm^{-3}
α	0.91	0.76	–

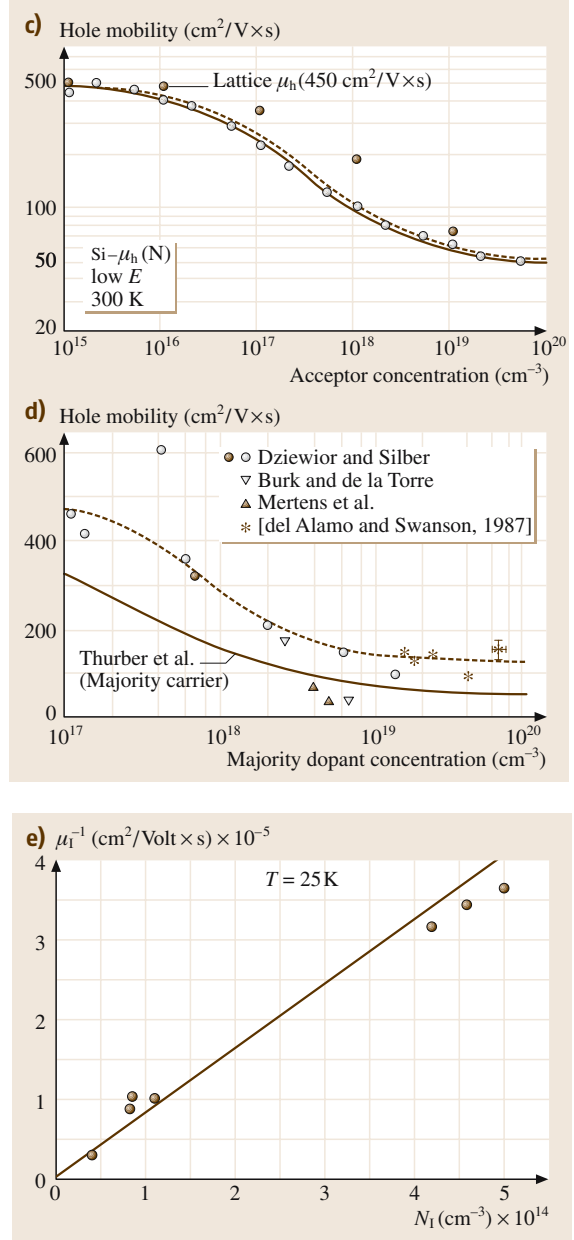


Fig. 21.8 (e) Inverse ionized-scattering mobility in n-Si at $T = 25 \text{ K}$ versus N_I . After [21.100] with permission

$T = 300 \text{ K}$ shows no visible variation when sweeping from low to high fields, which means that either the high-field regime is not attained or $r_e = 1$. The transition between the two regimes is clearly seen with $r_e < 1$, at the lower indicated values of T . The data below 20 K

Table 21.7 Best fit parameters for (21.26, 31, 32)

Parameters	Electrons	Holes	Units
μ_0	1400	480	cm^2/Vs
N_{ref}	3×10^{16}	4×10^{16}	cm^{-3}
S	350	81	...
E_1	3.5×10^3	6.1×10^3	V/cm
E_2	7.4×10^3	2.5×10^3	V/cm
F	8.8	1.6	...
v_m	$1.53 \times 10^9 \times T^{-0.87}$	$1.62 \times 10^8 \times T^{-0.87}$	cm/s
E_c	$1.01 \times T^{1.55}$	$1.01 \times T^{1.68}$	V/cm
β	$2.57 \times 10^{-2} \times T^{0.66}$	$0.46 \times T^{0.17}$...

Table 21.8 Intrinsic mobility in crystalline Si at room temperature

Carriers	Hall mobility ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$)			Drift mobility ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$)			
Electrons	1610 ^a	1450 ^b	1560 ^c	1500 ^d	1610 ^a	1350 ^e	1360 ^c
Holes	365	298	345	500	360	480	510

^a [21.95], ^b [21.96], ^c [21.94], ^d [21.92], ^e [21.93]

Table 21.9 Room-temperature mobility of Si at $n(p) = 2 \times 10^{18} \text{ cm}^{-3}$ [21.112]

Impurity	Donor		Acceptor	
	Sb	As	B	Ga
Ionization energy (eV)	0.039	0.049	0.045	0.065
Mobility ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$)	235	220	110	100

Table 21.10 Si samples [21.96]

Sample number	Impurity	E_d or E_a (eV)	$N(\text{cm}^{-3})$	N_a or $N_d (\text{cm}^{-3})$	m^*/m_0
n-type					
131	As	0.056	1.75×10^{14}	1.00×10^{14}	0.5
130	As	0.049	2.10×10^{15}	5.25×10^{14}	1.0
129	As	0.048	1.75×10^{16}	1.48×10^{15}	1.2
139	As	0.046	1.30×10^{17}	2.20×10^{15}	1.0
126	As	?	2.00×10^{18}
140	As	Degenerate	2.70×10^{19}
p-type					
159	B	0.045	3.10×10^{14}	4.10×10^{14}	0.4
127	B	0.045	7.00×10^{14}	2.20×10^{14}	0.4
117	B	0.043	2.40×10^{16}	2.30×10^{15}	0.6
119	B	0.043	2.00×10^{17}	4.90×10^{15}	0.7
141	B	?	1.00×10^{18}
125	B	Degenerate	1.50×10^{19}

Table 21.11 Si samples [21.100]

Sample	Impurity	$N_d (\text{cm}^{-3})$	$N_a (\text{cm}^{-3})$	$N_{\text{Oxygen}} (\text{cm}^{-3})$
A (FZ)	P	1.14×10^{14}	4.00×10^{12}	$\approx 10^{16}$
B (FZ)	P	9.00×10^{13}	2.00×10^{13}	$\approx 10^{16}$
C (CZ)	P	2.50×10^{14}	5.50×10^{13}	$\approx 10^{18}$
D (CZ)	P	4.90×10^{14}	2.10×10^{14}	5.0×10^{17}
E (CZ)	P	3.30×10^{14}	2.30×10^{14}	7.7×10^{17}
F (FZ)	B	2.00×10^{12}	3.40×10^{14}	$\approx 10^{16}$

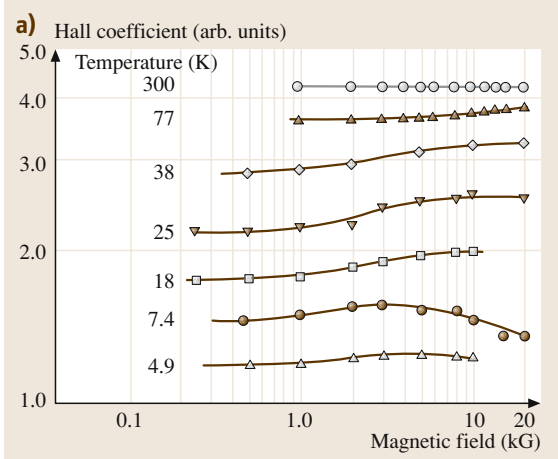


Fig. 21.9 (a) Hall-effect parameters in Si: (a) Hall coefficient in n-Si versus magnetic field strength as measured for Si:AsI (Table 21.13) at several indicated temperatures. After [21.53] with permission;

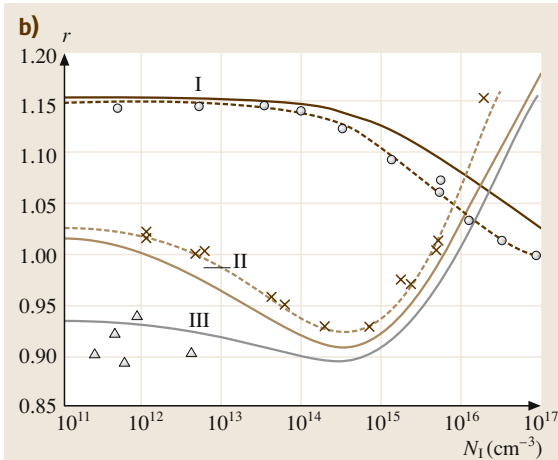


Fig. 21.9 (b) Hall factor in n-Si versus N_I , cross and open circle – experimental data at $T = 77$ K and $T = 300$ K, respectively, open triangle – experimental data at $T = 77$ K for samples with lowest μ_e (supposedly compensated), dashed lines – guides for the eye, full lines are the dependencies calculated using theory [21.103] at $T = 77$ K (I) and 300 K (II) without account of neutral-impurity scattering, and at $T = 77$ K on account of neutral impurities with the density $6 \times 10^{16} \text{ cm}^{-3}$ (III). After [21.106] with permission;

were obtained by sweeps of the photo-Hall coefficient, so the decrease of R_e^H at high B is fictitious, due to the field dependence of the photo-carriers lifetime [21.53].

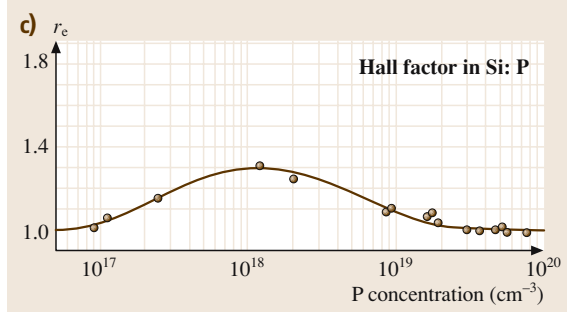


Fig. 21.9 (c) Hall factor in highly doped n-Si:P at $T = 300$ K versus phosphorous concentration [21.105];

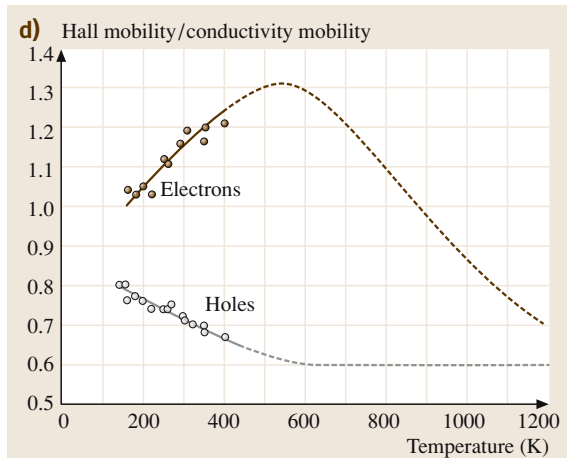


Fig. 21.9 (d) Hall factors for electrons and holes versus T , solid circle – measured, dashed line – computed dependencies. After [21.96] with permission;

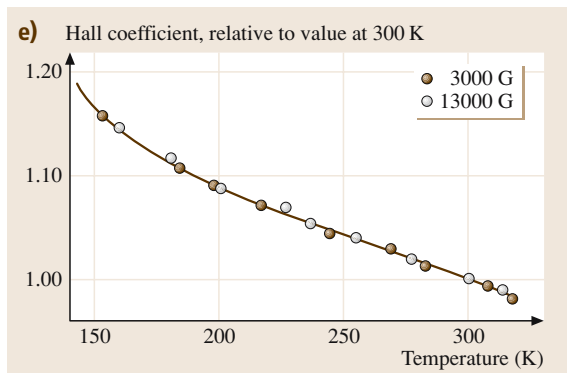


Fig. 21.9 (e) Hall coefficient, relative to its value at 300 K, versus T at two indicated strengths of magnetic field in a sample of p-Si with room-temperature $\rho = 35 \Omega\text{cm}$. After [21.98] with permission;

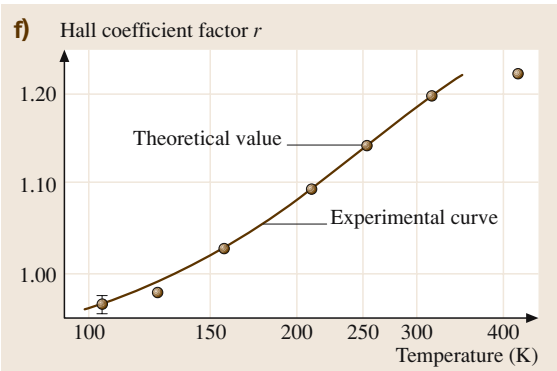


Fig. 21.9 (f) Electron Hall factor versus T in n-type (SP6A, Table 21.12) sample, *full line* – experimental data, *open circle* – theory. After [21.51] with permission

Figures 21.9b and 21.9c display the dependence of r_e on the impurity concentration for light to moderate and high doping, respectively. Figures 21.9a and 21.9b are in qualitative agreement with the inequality $r_e < 1$ at $T = 77$ K for comparable concentrations. The overall observed behavior of $r_e(N)$ at $T = 300$ K, as shown in Fig. 21.9c,

Table 21.12 Compensated Si:P samples [21.51,66]

Sample	N_d (cm ⁻³)	N_a (cm ⁻³)
SP6X	1.7×10^{13}	8.0×10^{12}
SP6S	2.1×10^{13}	8.0×10^{12}
SP6A	2.8×10^{13}	7.5×10^{12}
SP8A	5.4×10^{13}	1.0×10^{13}
SP4M	1.3×10^{14}	1.0×10^{13}
SP4A	2.4×10^{14}	1.0×10^{13}
SP1A	6.7×10^{14}	3.8×10^{14}
SP1D	8.0×10^{14}	4.0×10^{14}
SP2A	2.3×10^{15}	8.3×10^{14}
SM 2	3.9×10^{15}	3.3×10^{15}
SM 3	4.5×10^{15}	3.3×10^{15}

Table 21.13 Weakly compensated n-Si samples [21.53]

Sample	N_d (cm ⁻³)	N_a (cm ⁻³)	E_d (meV)	m^*/m_0	$\rho_{300}(\Omega \cdot \text{cm})$
Si:P 1	9.5×10^{15}	4.2×10^{12}	45.64	0.3218	0.66
Si:P 3	2.5×10^{14}	2.3×10^{13}	45.14	0.3218	17.1
Si:P 4	9.6×10^{14}	2.0×10^{14}	43.39	0.3218	4.76
Si:P 5	3.3×10^{14}	8.6×10^{12}	45.38	0.3218	13.6
Si:P 6	4.3×10^{13}	7.7×10^{12}	45.42	0.3218	123
Si:As 1	1.4×10^{16}	6.0×10^{12}	53.64	0.3218	0.43
Si:As 2	7.9×10^{15}	4.3×10^{13}	52.54	0.3218	0.66
Si:As 3	7.5×10^{16}	1.8×10^{13}	52.52	0.3218	0.13
Si:Sb 1	7.4×10^{15}	5.3×10^{12}	42.59	0.3218	0.68

viz. $r_e \approx 1$ on both the moderate and heavy doping sides and a maximum in between at which $r_e > 1$, is deduced from (21.23) using knowledge of the anisotropy of the mass and relaxation-time tensors. Compared to these, results for the Hall effect for holes are not available, to our knowledge.

Hall Effect and Mobility Versus Temperature
The dependencies of $r_e(T)$ and $r_h(T)$ for samples 131, 130 and 159, 127 from Table 21.10 are shown in Fig. 21.9d. The temperature dependence of R_h^H relative to its value at $T = 300$ K for another p-type sample is shown in Fig. 21.9e. Since, in the considered temperature range, p is constant with T , the dependence is fully congruent to $r_h(T)$ – in this regard the lower curve in Fig. 21.9d and the curve in Fig. 21.9e are in agreement. There is good quantitative agreement between experiment and theory for $r_e(T)$ using an exact effective-mass tensor, as shown in Fig. 21.9f. On the contrary, the Hall effect and MR in p-Si are poorly understood [21.98,99] as the energy spectra of holes is much more complicated than (21.5, 6), even for energies well above 200 K [21.7].

The $\mu_{e,h}(T)$ curves have been the subject of extensive, but to our opinion not exhaustive, studies. As discussed in Sect. 21.2, these dependencies reflect the scattering mechanisms. The basic question is that of temperature dependence due to lattice scattering, which can be measured on technologically pure, very lightly doped and negligibly compensated samples. Since such samples were not available in early studies, researchers tried to apportion the intrinsic $\mu_{e,h}(T)$ from the drift measurements, as shown in Fig. 21.10a, or from conductivity and Hall measurements combined with calculations, as shown in Fig. 21.10b. In practice the intrinsic $\mu_{e,h}(T)$ is described by the phenomenological best-fit equation

$$\mu_{e,hL}(T) = AT^{-\gamma} \quad , \quad (21.29)$$

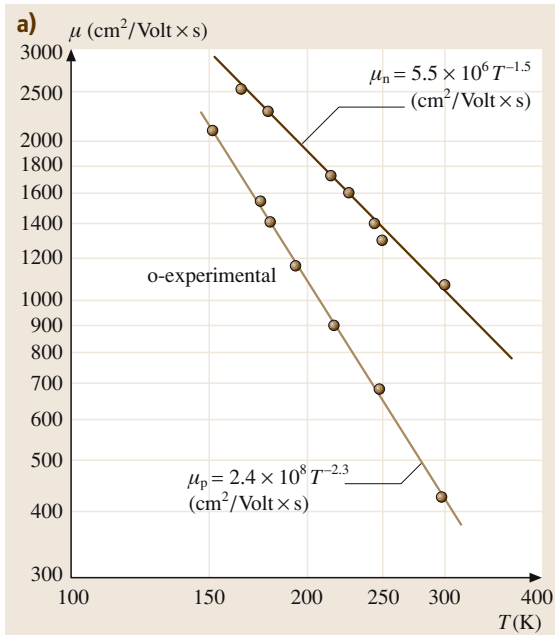


Fig. 21.10 (a) Apportionment of dependence of lattice mobility on T in Si: (a) Minority electron and hole drift mobility in high-resistivity crystals. After [21.92] with permission;

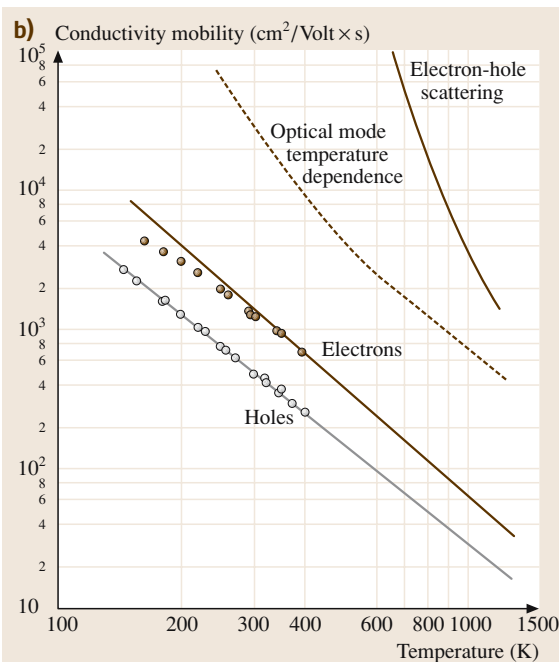


Fig. 21.10 (b) Conductivity mobility in TL crystals of n-type (130, 131, Table 21.10) and p-type (127, 159, Table 21.10). After [21.96] with permission;

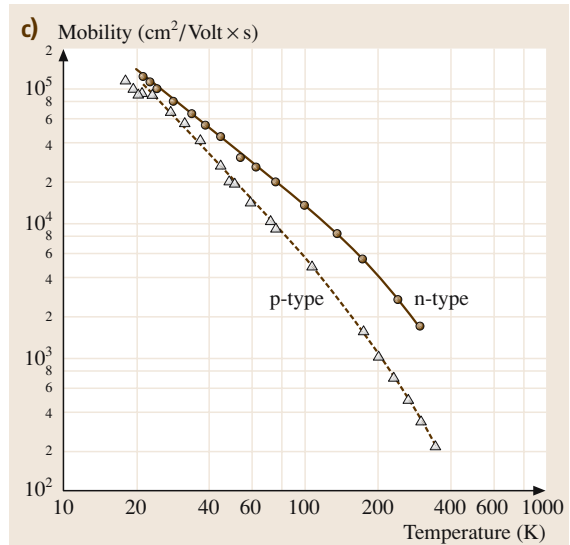


Fig. 21.10 (c) Conductivity mobility in FZ crystals of n-type (A, Table 21.11) and of p-type (F, Table 21.11). After [21.100] with permission;

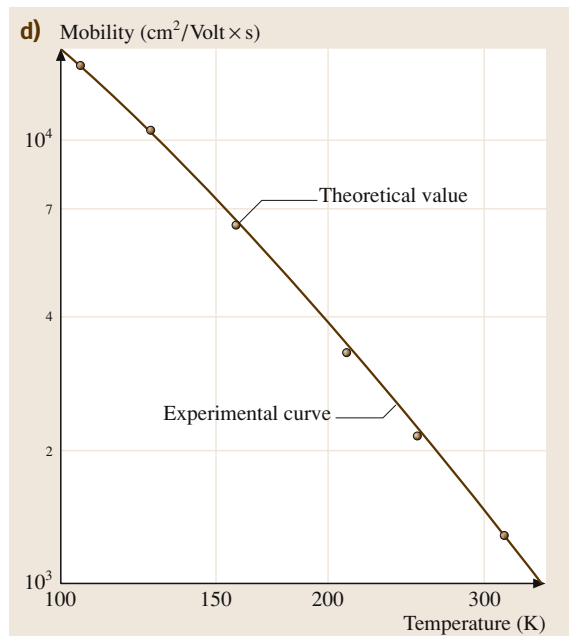


Fig. 21.10 (d) Conductivity mobility in FZ crystal of n-type (SP6A, Table 21.12), full line – experimental data, open circle – theory. After [21.51] with permission;

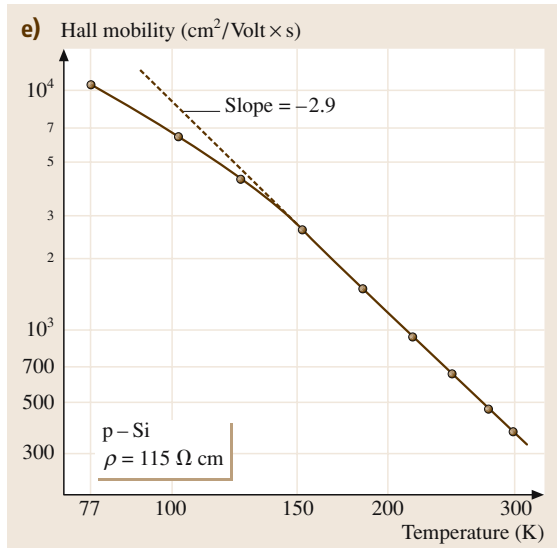


Fig. 21.10 (e) Hall mobility in a CZ crystal of p-type with the indicated resistivity. After [21.99] with permission

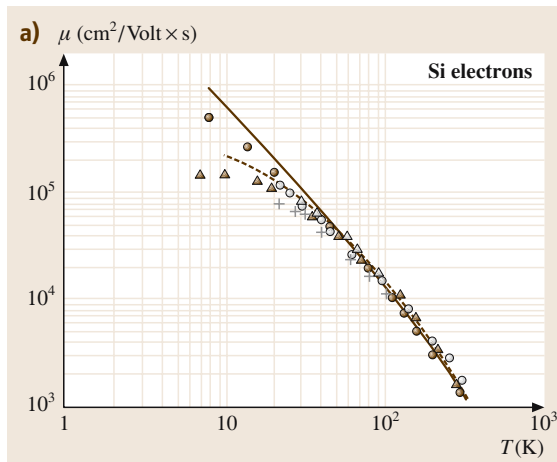


Fig. 21.11 Summary of data on μ versus T in nearly intrinsic to lightly doped Si; symbols and lines indicate experimental data and simulations, respectively: (a) n-type. After [21.114] with permission. Solid circle – time-of-flight technique on high-purity n-Si (Ohm's regime achieved only at $T \geq 45$ K) by Canali et al., plus – [21.115], open triangle – [21.101], open circle – [21.100], solid triangle – [21.53], dashed and full lines – Monte Carlo method by Canali et al. using three allowed and six lowest inter-valley phonon modes (Table 21.4), respectively, and no impurities, dash-dotted line – the same as full, but with inclusion of 10^{13} cm^{-3} ionized donors;

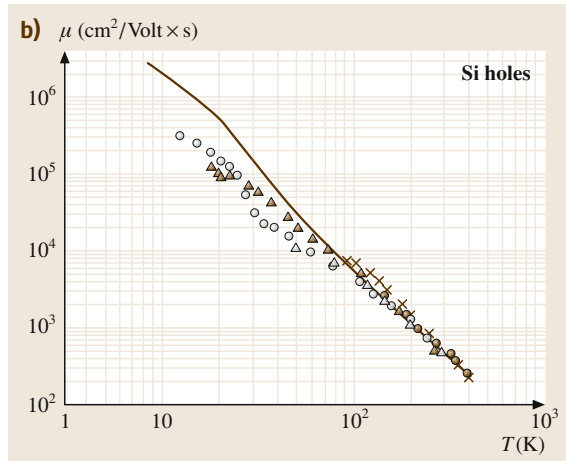


Fig. 21.11 (b) p-type. After [21.13] with permission. Open circle – time-of-flight technique on high-purity p-Si (Ohm's regime achieved only at $T \geq 100$ K) by Ottaviani et al., solid triangle – [21.100], cross – minority hole drift mobility [21.93], open triangle – [21.96], solid circle – [21.96], full line – relaxation time method by Ottaviani et al. with the use of simplified models of the hole energy spectrum and hole-phonon interactions

where A and γ may vary with the carrier and measurement type, fitting interval and sample quality. Different drift measurements, in the ranges 150–300 K [21.92], and 160–400 K [21.93], resulted in very different values of A and γ for both electrons and holes. The first measurements of the Hall effect and conductivity against temperature [21.96] gave the same γ for holes as that obtained by Prince, but a drastically larger γ for electrons, as the n-type samples with the lowest N_d used were inappropriate for that purpose – one sample (131, Table 21.10), though lightly doped, was compensated ($K \approx 0.57$) and the other (130, Table 21.10) was doped in excess of 10^{15} cm^{-3} . Therefore, Morin and Maita numerically extracted the ionized-impurity scattering mobility portion (21.27, 28) and then determined γ . This procedure seems unreliable in view of the inapplicability of the BH formula with strong compensation, neglect of scattering by neutral impurities and of the spectrum and scattering anisotropies. Analogously, using (21.29) for a high-resistivity p-type Si sample [21.99], fitting to $\mu_h(T)$ resulted in the same γ as that obtained by Ludwig and Waters while fitting to $\mu_h^H(T)$ (Fig. 21.10e) resulted in an even larger γ . Using high-purity samples, for which the $\mu_e(T)$ and $\mu_h(T)$ values are presented in Fig. 21.10c, made it possible to determine a γ for $\mu_e(T)$ in agreement with that obtained by Prince, but failed to fit $\mu_h(T)$

Table 21.14 Parameters of the best fit of the intrinsic mobility to (21.29)

Carriers	γ						$A(10^8 \text{ K}^\gamma \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1})$			
Electrons	1.5 ^a	2.6 ^b	2.5 ^c	...	1.5 ^e	2.42 ^f	5.5 ^a	40 ^b	21 ^c	14.3 ^f
Holes	2.3 ^a	2.3 ^b	2.7 ^c	2.7–2.9 ^d	2.0 ^e	2.20 ^f	2.4 ^a	2.5 ^b	23 ^c	1.35 ^f

^a [21.92], ^b [21.96], ^c [21.93], ^d [21.99], ^e [21.100] at $T \leq 100 \text{ K}$, ^f [21.14]

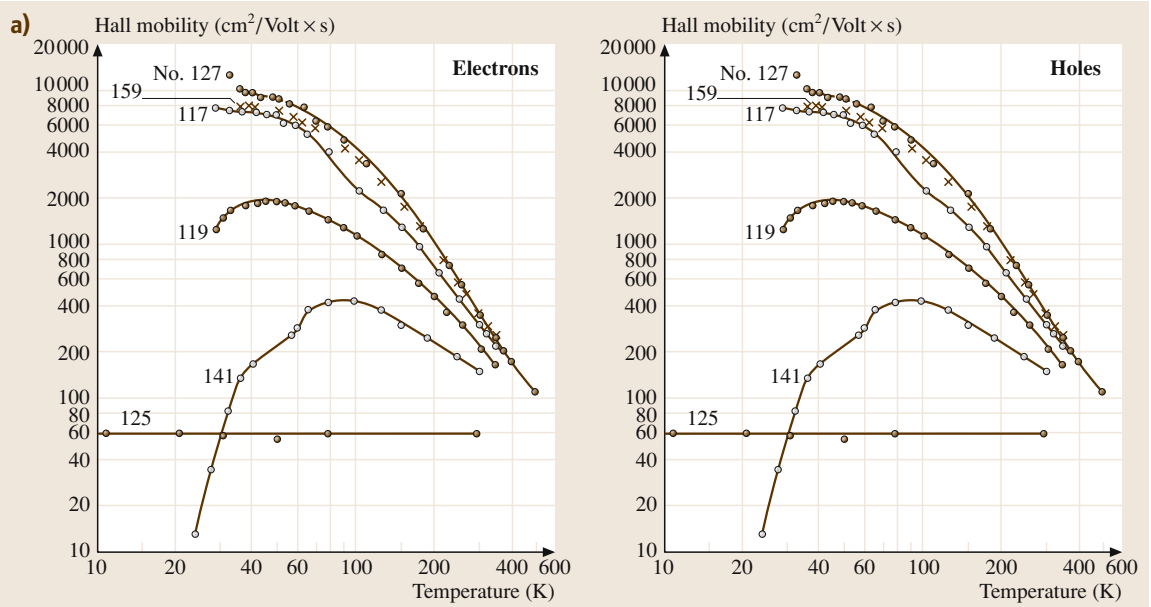


Fig. 21.12 Effects of doping on mobility versus T curves in Si: **(a)** Conductivity mobility of differently doped n- and p-type TL crystals. Numbers indicate samples from Table 21.10. After [21.96] with permission;

to (21.29) [21.100]. Long did not fit using (21.29) and calculated $\mu_{eL}(T)$ using a realistic electron spectrum and a specific inter-valley electron–phonon scattering model (Sect. 21.1.4TSnotePlease check this link.), and compared the calculation to experiment [21.51], with excellent agreement, as can be seen in Fig. 21.10d. The validity of the model [21.51] was confirmed by an independent study [21.53]. The calculation of $\mu_{eL}(T)$ was revisited on the basis of the Monte Carlo method and extended inter-valley scattering models [21.113]; the results are displayed in Fig. 21.11a along with the experimental data and fit line. We are unaware of analysis for $\mu_{hL}(T)$ comparable to that done for $\mu_{eL}(T)$, taking both the realistic hole spectrum [21.7] and full hole–phonon scattering into consideration. Another analysis [21.70] greatly simplified the problem by using relaxation-time equalization ($\tau_1(E) = \tau_2(E)$), resulting from the isotropic approximation adopted for the hole spectra (21.5), but proved insufficient to account for both Ohm’s mobility and hot-holes phenomena [21.13]. The results

of calculations based on the relaxation-time method, using only one isotropic but nonparabolic heavy-hole band [21.13], are displayed in Fig. 21.11b along with the experimental data and fit line. The best-fit parameters of (21.29) obtained by different authors are given in Table 21.14.

The effect of doping on mobility is shown in Fig. 21.12. In doped samples $\mu_{e,h}(T)$ increases with decreasing T up a maximum at some T_{\max} (the lower the sample’s purity the larger T_{\max}) and then starts to decrease. This behavior is due to competition between lattice scattering, which weakens with decreasing T , and impurity scattering, which strengthens with decreasing T , (21.27, 28, 29). In heavily doped samples scattering by impurities, while dominating the mobility, ceases to depend notably on T [(21.27, 28, 29) do not hold in this case], which results in an overall weak dependence of the observed $\mu_{e,h}(T)$ on T . These types of behavior are clearly seen in Fig. 21.12a–b. The evolution of T_{\max} with the impurity content for lightly doped, differ-

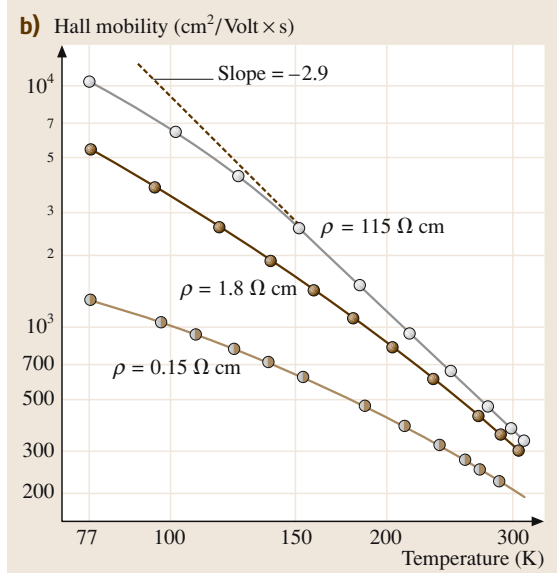


Fig. 21.12 (b) Hall mobility of p-type CZ crystals with different ρ_{300} . After [21.99] with permission;

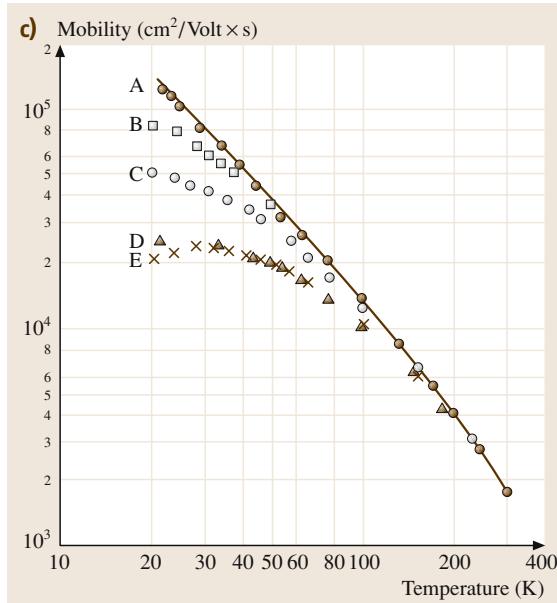


Fig. 21.12 (c) Conductivity mobility versus T of lightly doped, differently compensated, n-Si samples from Table 21.11. After [21.100] with permission;

ently compensated, n-Si is readily traced in Fig. 21.12c. The compensation is crucial for the rate of downturn of the mobility, as follows from (21.27, 28) and is confirmed by the curves for samples C and D (Table 21.11)

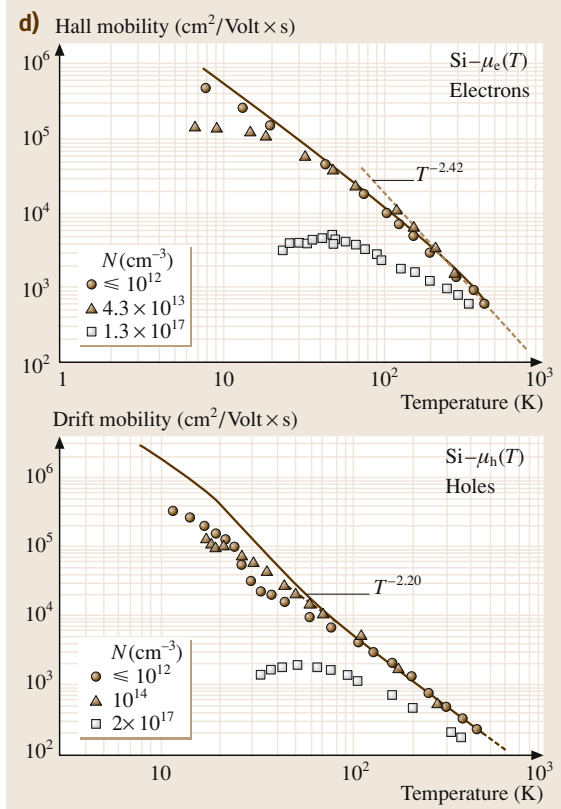


Fig. 21.12 (d) Summary of data on μ versus T ; symbols and lines indicate experimental data and simulations, respectively. After [21.14] with permission. Electrons: solid circle – high purity Si ($N \leq 10^{12} \text{ cm}^{-3}$, time-of-flight) [21.114], solid triangle – lightly doped compensated (Si: P 6, Table 21.13, photo-Hall effect) [21.53], open square – moderately doped with $K \approx 0.01$ (139, Table 21.10, Hall effect) [21.96], full line indicates the theoretical results for the lattice mobility [21.114], dot-dashed line gives the best fit of that mobility by an inverse power of T around room temperature ((21.29), Table 21.9); Holes: solid circle – high purity Si ($N \leq 10^{12} \text{ cm}^{-3}$, time-of-flight) [21.13], open triangle – lightly doped with $K \approx 0.01$ (F , Table 21.11, Hall effect) [21.100], open square – moderately doped with $K \approx 0.1$ (119, Table 21.10, Hall effect) [21.96], full line indicates the theoretical results for the lattice mobility, dot-dashed line gives the best fit of that mobility by an inverse power of T around room temperature ((21.29), Table 21.9) [21.13]

in Fig. 21.12c. Figure 21.12d summarizes the information on $\mu_{e,h}(T)$ for doped Si. For estimations of the

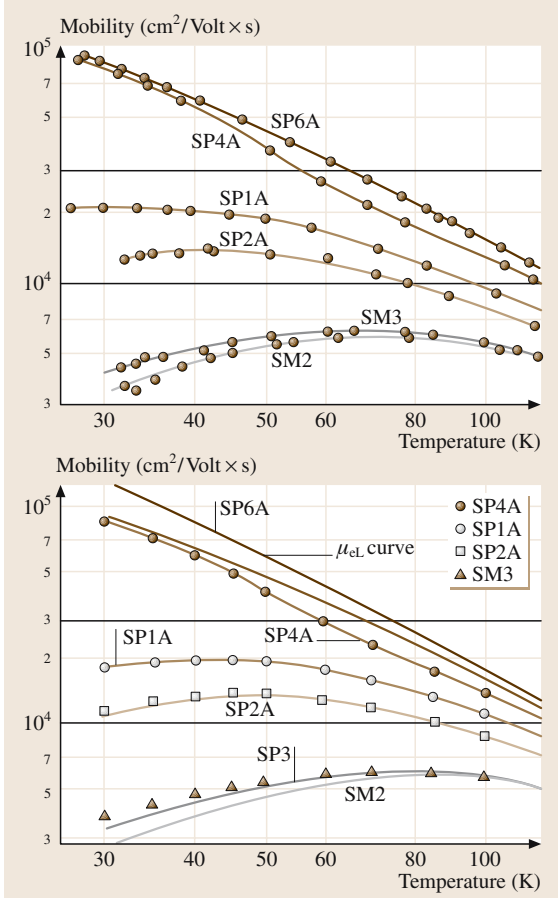


Fig. 21.13 Comparison of the theoretical results on $\mu_e(T)$ and $\mu_e^H(T)$ (symbols) with the experimental data (full lines). The indicated samples are those from Table 21.12. After [21.101] with permission

combined-scattering mobility e.g. for electrons, the Mat-tisen rule $\mu_e^{-1}(T) = \mu_{eL}^{-1} + \mu_{eI}^{-1}$, where μ_{eL} and μ_{eI} are given by (21.29) and (21.27, 28), respectively, is used (e.g. [21.100, 101]). The formula, consistently derived for nondegenerate carriers in the standard band, is

$$\mu_e(T) = \mu_{eL} + \mu_{eL}x^2 \times \left\{ \text{Ci}(x) \cos x + \left[\text{Si}(x) - \frac{\pi}{2} \right] \sin x \right\},$$

$$x = \sqrt{\frac{6\mu_{eL}}{\mu_{eI}}}, \quad (21.30)$$

where $\text{Ci}(x)$ and $\text{Si}(x)$ are integral cosine and sine, respectively, and μ_{eL} is due to acoustic-phonon scattering (with $\gamma = 1.5$) [21.71], which proved more appropri-

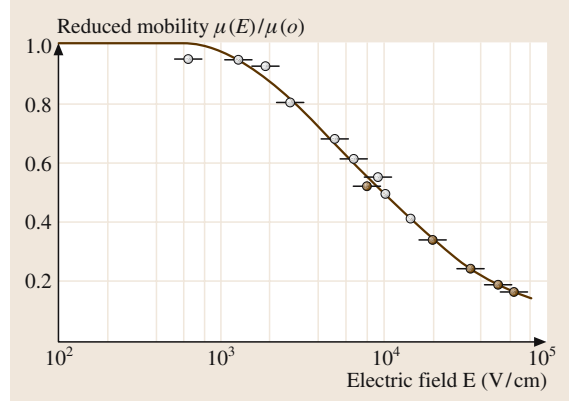


Fig. 21.14 Variation of reduced mobility with electric field applied to a 70 Ωcm n-Si sample along the direction (111). The electron concentration is assumed to be constant. After [21.24] with permission

ate for quantitative analysis. The comparison of (21.30) and the expression for $\mu_e^H(T)$, derived under the same conditions, with the corresponding experimental curves, using $\eta = 8.6 \times 10^{17} \text{ K}^{-1.5} \text{ cm}^2 \text{ V}^{-2} \text{ s}^{-1}$ (21.27) for an effective isotropic band mass of $m^* = 0.3m_0$, is shown in Fig. 21.13. Good quantitative agreement between the theory and experiment is observed except for the samples SM3 and SM2, in which the ionized-impurity scattering is strongest (to avoid confusion the data points for SM2 are not shown but they are as discrepant with respect to the observed mobility curve as those shown for SM3). The generalization of (21.30) to the case with spectrum and scattering anisotropies [21.103] was employed to calculate $r_e(T)$ [21.65] and to extract N_a from the mobility data [21.53]. Dakhoskii numerically explained why the above isotropic approximation [21.101] worked well, but presented no graphs of mobility versus T . Norton et al. presented graphically calculated $\mu_e(T)$ only for Si:P6, the purest of their samples. No analysis for $\mu_h(N, T)$ comparable to that done for $\mu_e(N, T)$, i.e. taking both the realistic hole spectrum and hole-phonon and hole-impurity scatterings into account, exists to our knowledge in the literature. Success of modeling based on empirically adjusted phenomenological expressions (point 2–4 in Sect. 21.2.3), though valuable for the device community, cannot be considered as such from the viewpoint of solid-state and semiconductors theories.

21.2.2 High-Electric-Field Effects

The group of effects associated with the carrier transport in high electric fields is traditionally called

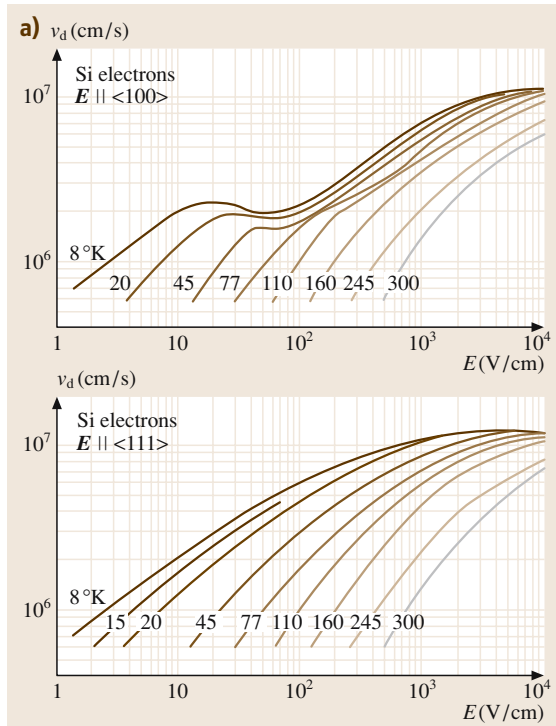


Fig. 21.15 Carriers' drift velocity for E , applied along two indicated directions, measured at different indicated T by the time-of-flight method: **(a)** Electrons. After [21.116] with permission;

hot-carrier phenomena. This name was borrowed from the displaced-Maxwell-distribution method, where an effective parameter is the carrier temperature, which becomes much higher than that of the lattice, as if the field heats the carriers [21.38]. Not too much work has been published on hot-carrier phenomena in Si, but most of the existing work was overviewed in the two references cited below, and in point 3 in Sect. 21.2.3.

1. The electron drift velocity ([21.113]). Experimental results for electrons in Si obtained using a time-of-flight technique were presented for $8\text{ K} \leq T \leq 300\text{ K}$ and E in the magnitude range $1.5 \times 10^4 - 5 \times 10^4\text{ V/cm}$, oriented along the $\langle 111 \rangle$, $\langle 110 \rangle$ and $\langle 100 \rangle$ crystallographic directions. At 8 K the dependence of the transit time upon sample thickness allowed a measurement of the valley repopulation time when the electric field is $\langle 100 \rangle$ oriented. These experimental results were interpreted with Monte Carlo calculations in the same ranges of T and E . The model

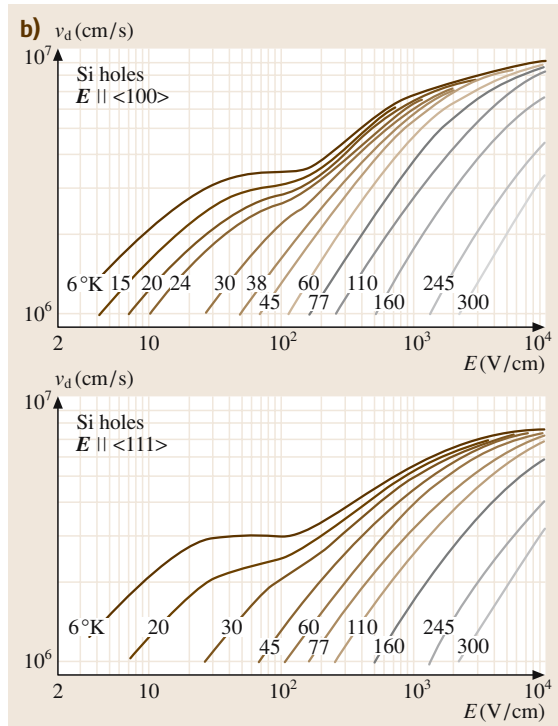


Fig. 21.15 (b) Holes. After [21.13] with permission

included the many-valley structure of the conduction band of Si, acoustic intra-valley scattering with the correct momentum and energy relaxation and correct equilibrium phonon population, several inter-valley scatterings, and ionized-impurity scattering.

2. Hole drift velocity [21.13]. Drift velocities for holes in high-purity Si, were measured by a time-of-flight technique with E in the amplitude range from $3 \times 10^4 - 5 \times 10^4\text{ V/cm}$ along the $\langle 100 \rangle$, $\langle 110 \rangle$, and $\langle 111 \rangle$ directions, and at $8\text{ K} \leq T \leq 300\text{ K}$. The Ohm's mobility is theoretically interpreted on the basis of a two-band model consisting of a spherical parabolic and a spherical nonparabolic band, and relaxation-time approximation. The low- T Ohm's mobility proved to be strongly influenced by the nonparabolic structure of the heavy-hole band. The high-field region, $E = |E| \geq 103\text{ V/cm}$, was analyzed using a single, warped heavy-hole band model and a Monte Carlo technique. Anisotropy of the hot-hole drift velocity due to warping of the valence band was measured. Optical- and acoustic-phonon scattering were found to have comparable rates.

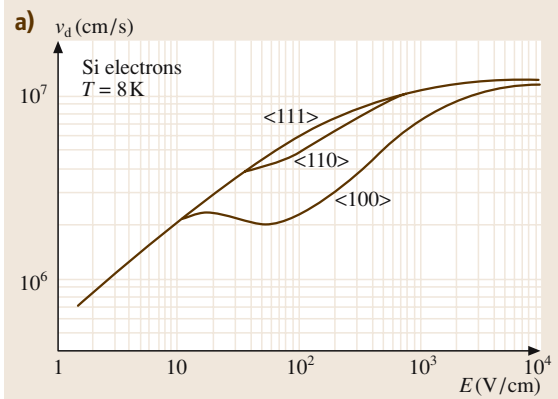


Fig. 21.16 Carriers' drift velocity anisotropy in details at different indicated T : (a) Electrons. After [21.116] with permission;

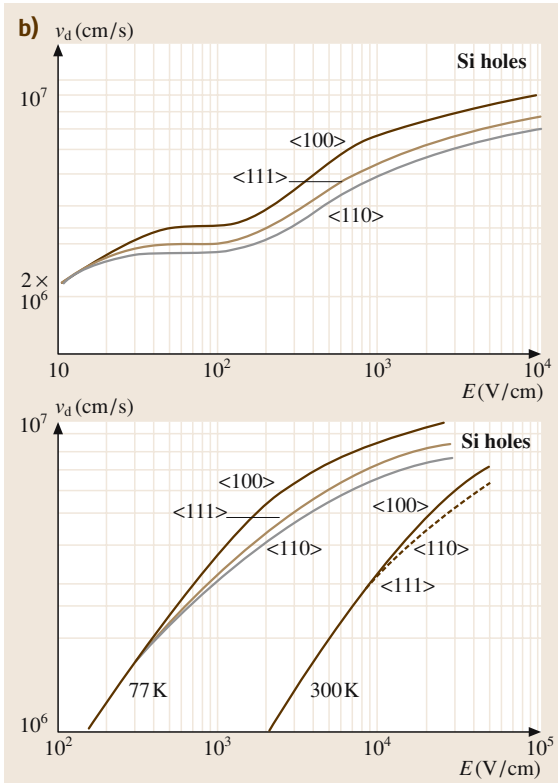


Fig. 21.16 (b) Holes. After [21.13] with permission;

Impact ionization phenomenon are different in appearance and are not considered here in detail; see point 7 in Sect. 21.2.3.

Drift Velocity – Electric-Field Relationship

As the field strength E overcomes some limit (which depends on the lattice temperature, doping and type of carrier) the drift velocity (Sect. 21.2.3) becomes a sub-linear function of E , i.e. the mobility starts to decrease with increasing E . As revealed by the early study for n-Si [21.117], in the range 10^3 – 10^5 V/cm, n does remain constant, but the mobility decreases appreciably, as seen Fig. 21.14. Further studies obtained, in addition, the following patterns: (e.g. [21.13, 14, 113]).

1. Drift velocity $v_d(E)$ is anisotropic in the orientation of E relative to the crystallographic axes: different values of v_d are obtained for the same E for E along different high-symmetry directions $\langle 111 \rangle$, $\langle 110 \rangle$ and $\langle 100 \rangle$. For less-symmetric directions v_d is not even parallel to E , which is known as the Sasaki–Shibuya effect [21.38]. The v_d – E relation depends strongly on T , as seen from Fig. 21.15a–b for the $\langle 111 \rangle$ and $\langle 100 \rangle$ directions. The anisotropy becomes stronger as T decreases, as seen from Fig. 21.16a–b. For electrons, the curves of v_d versus E for different T (Fig. 21.15a) tend to join together at some value of E , which increases with T , even though such a rejoining is not reached at all considered T . For holes (Fig. 21.15b), this tendency is much weaker and the rejoining field is much higher than for electrons.
2. For electrons at the highest $E \parallel \langle 111 \rangle$, a region where v_d is independent of E (saturation) is obtained (Fig. 21.15a). For holes, this type of saturation is not approached for all available E (Fig. 21.15b).
3. Ohm's law is reached at $T > 45$ K and 100 K for electrons and holes, respectively, and the results for mobility are in agreement with those obtained by standard techniques for the linear regime. At $T < 45$ K for electrons, and $T < 100$ K for holes, even at the lowest E , v_d is not linear with E . At low temperatures the mobility is as high as $5 \times 10^5 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ at $E = 1.5 \text{ V/cm}$ for electrons ($T = 8 \text{ K}$), and $3.5 \times 10^5 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ at $E = 3 \text{ V/cm}$ for holes ($T = 6 \text{ K}$). A negative differential mobility (NDM) was found with $E \parallel \langle 100 \rangle$ at $T < 40 \text{ K}$ for electrons (Fig. 21.16a). The field at which v_d reaches the maximum before the NDM region (NDM threshold), decreases with decreasing T . For holes, a net low-field saturation region of v_d ($50 \text{ V/cm} < E < 150 \text{ V/cm}$ at $T = 6 \text{ K}$) shows up for $E \parallel \langle 111 \rangle$, $\langle 110 \rangle$ and $\langle 100 \rangle$ (see top of Fig. 21.16b). This effect gradually disappears as T increases, and at $T > 38 \text{ K}$ it is no longer visible.

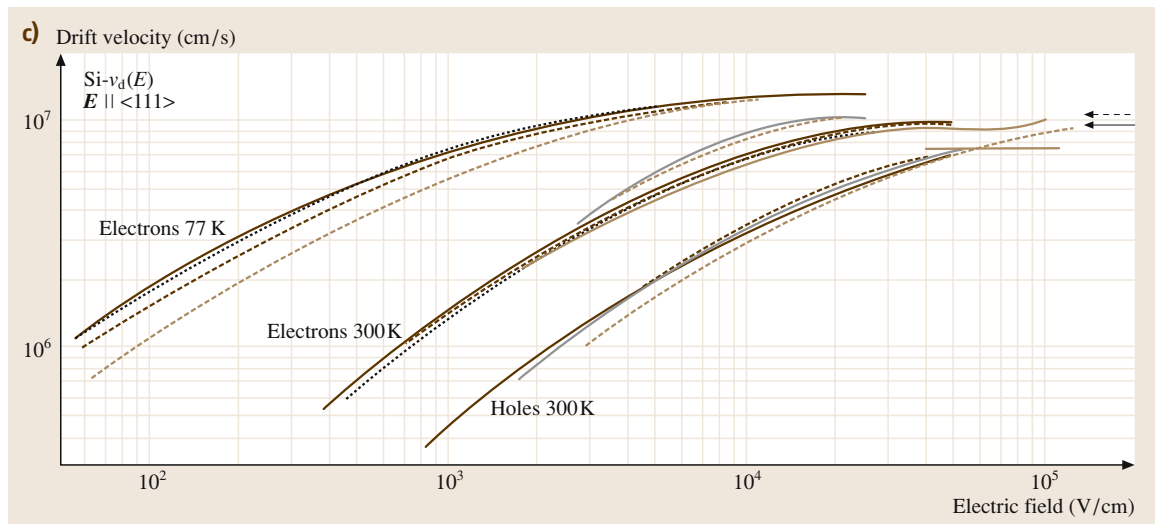


Fig. 21.16 (c) Summary of the data on the drift velocity obtained by different techniques. After [21.14] with permission. Holes, $T = 300$ K: *full* – time-of-flight [21.118], *dashed* – time-of-flight [21.119], *double-dot-dashed* – time-of-flight [21.120], *cross-dashed* – $I(V)$ in space-charge-limited current (SCLC) regime [21.121], *dot-dashed* – $I(V)$ [21.122], *arrow* – extrapolated value of v_{sh} [21.122]. Electrons, $T = 300$ K: *full* – time-of-flight [21.118], *long-dash-dot* – time-of-flight [21.119], *dashed* – time-of-flight [21.120], *double-dot-dash* – $I(V)$ in the SCLC regime [21.123], *dots* – $I(V)$ [21.124], *dash-crossed* – $I(V)$ [21.125], *dashed arrow* – v_{se} from $I(V)$ in avalanche diodes [21.126, 127]. Electrons, $T = 77$ K: *full* – time-of-flight [21.118], *dashed* – $I(V)$ [21.128], *dash-dot* – $I(V)$ [21.129], *dotted* – $I(V)$ [21.130, 131]

For electrons, the anisotropy is due to a repopulation of the valleys: when $E \parallel \langle 111 \rangle$ the six valleys (Fig. 21.3b) are equally oriented with respect to E and all of them give the same contribution to v_d . When, for example, $E \parallel \langle 100 \rangle$, two valleys exhibit the effective mass m_l in the direction of the field, while the remaining four exhibit the transverse mass $m_t < m_l$. Electrons in the transverse valleys respond with a higher mobility, are heated to a greater extent by the field and transfer electrons to the two longitudinal, *colder* and slower valleys, which results in a lower v_d than for $E \parallel \langle 111 \rangle$, as seen e.g. in Fig. 21.16a. For holes, the anisotropy is due to the two warped and degenerate valence sub-bands (Fig. 21.3c) resulting in different effective masses for holes with different k , and in the lowest and highest v_d for $E \parallel \langle 110 \rangle$ and $E \parallel \langle 100 \rangle$, as seen in Fig. 21.16b. A summary of the curves of v_d versus E is presented on Fig. 21.16c.

Anisotropy becomes stronger as T decreases, since a lower T leads to less-effective relaxation effects. In particular, at $T \approx 45$ K the repopulation of electron valleys may be so rapid with increasing $E \parallel \langle 100 \rangle$ that NDM occurs with E in the range 20–60 V/cm. NDM was observed via oscillations of the current [21.132], and in the I – V [21.128, 132, 133] and v_d – E ([21.113, 134, 135]) characteristics. Whether the

curves of electron v_d versus E for $E \parallel \langle 100 \rangle$ and $E \parallel \langle 111 \rangle$ join together at the high-field limit is still an open point. Theoretical considerations [21.136] seem to indicate that a small difference ($\approx 5\%$) between the two curves should remain. However, this difference is comparable to experimental error and merging of the two curves has been claimed at several temperatures in the experimental results [21.113, 118]. In the case of holes, both theoretical and experimental results [21.13] indicate that, up to the highest fields attained (5×10^4 V/cm), no merging occurs. The aforementioned anomalous behavior of the curves of v_d versus E for holes at low temperatures ($T < 38$ K) – that these curves tend to *saturate* at intermediate field strengths and then rise again – has been explained [21.13, 116] by a nonparabolic distortion of the heavy-hole spectrum.

A general tendency of $v_d(E)$ to saturate at the highest fields is an important phenomenon which was considered from the very beginning of the hot-electron investigations [21.137–139]. The original explanation [21.138] for this behavior was based on a rough physical model, in which it was assumed that the carriers emit an optical phonon as soon as they reach the energy $\hbar\omega_0$ of the phonon that limits the drift veloc-

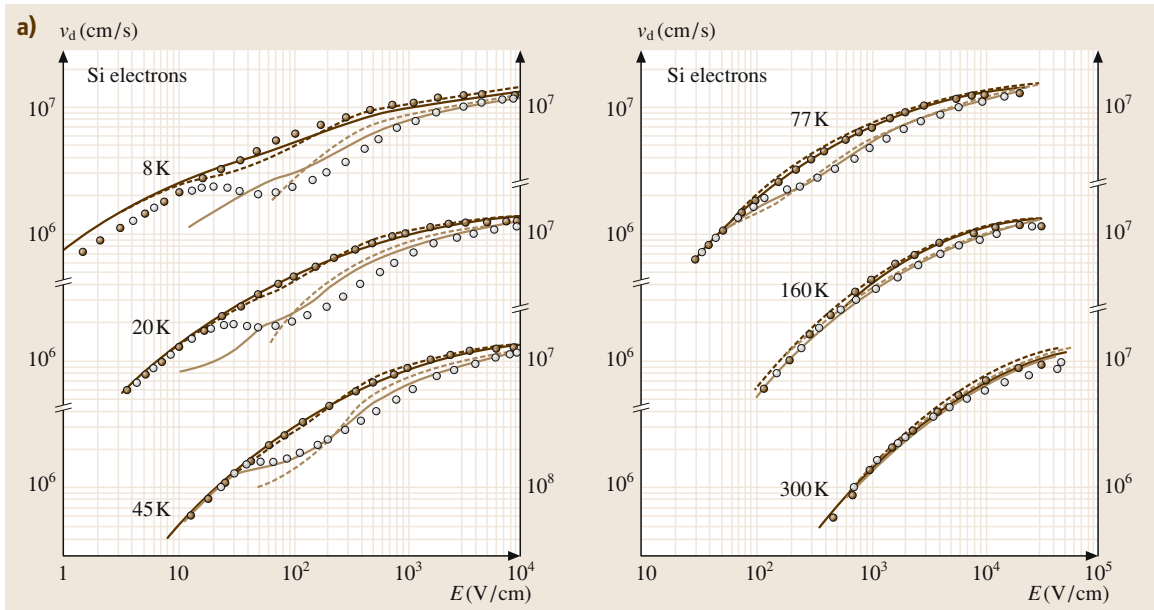


Fig. 21.17 Measured versus simulation results on drift velocity in Si: (a) $v_{de}(E)$ at different indicated temperatures. Circles indicate experimental data by time-of-flight method, solid circle – $E \parallel \langle 111 \rangle$ and $E \parallel \langle 100 \rangle$, and lines Monte Carlo simulations results neglecting impurity scattering, full and broken – using six lowest and three allowed inter-valley phonons (Table 21.4), respectively. For $T \leq 45$ K, the theoretical curves which refer to the $\langle 100 \rangle$ direction are interrupted, as it was not possible to reach a sufficient precision in the simulated v_d . After [21.113, 114] with permission;

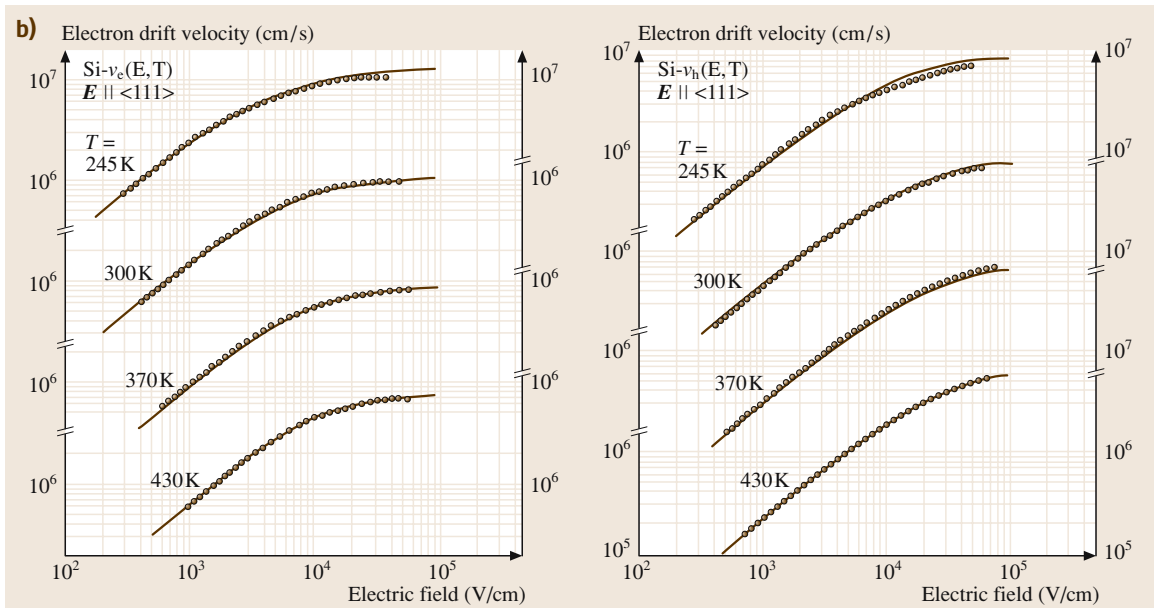


Fig. 21.17 (b) The $v_{de}(E)$ and $v_{dh}(E)$, for $E \parallel \langle 111 \rangle$, in high-purity Si at four different T . After [21.14] with permission, solid circle – the experimental data [21.118] and lines are the best-fit curves obtained with (21.32) using the parameters listed in Table 21.7;

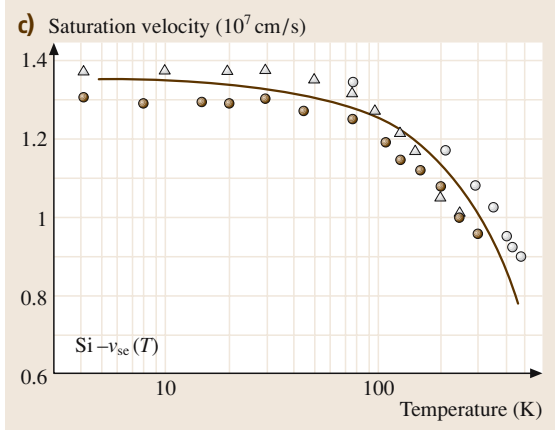


Fig. 21.17 (c) $v_{se}(T)$ obtained experimentally with different techniques and numerically fitted. After [21.14] with permission, *solid circle* – time-of-flight [21.113, 114, 118, 142], *open circle* – $I(V)$ in the avalanche diodes [21.126, 127], *open triangle* – $I(V)$ in the SCLC regime [21.123]; the latter data have been normalized to $9.6 \times 10^6 \text{ cm s}^{-1}$ at 300 K; *line* represents the best fitting curve of (21.33)

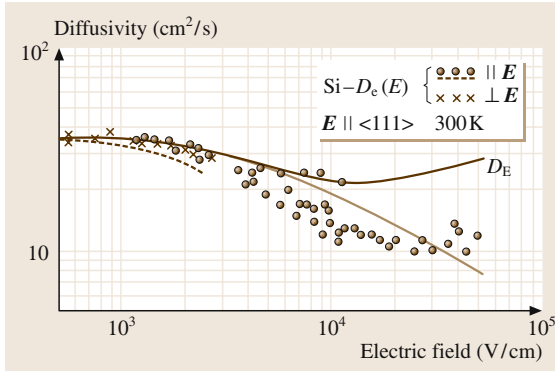


Fig. 21.18 Diffusion coefficients of electrons versus E in Si at $T = 300 \text{ K}$. After [21.14] with permission: *solid circle* – $D_{||}$, time-of-flight data [21.113, 114]; *dashed* – $D_{||}$, noise data [21.143]; *crosses* – D_{\perp} , the geometrical technique data [21.144]; *full lines* – the Monte Carlo computed $D_{||}$ and D_E [21.113, 114]

ity by the saturation value $v_s \approx \sqrt{\hbar\omega_0/m^*}$ [21.138]. For electrons, inter-valley instead of optical phonons should be considered (Sect. 21.2.3). Subsequent calculations [21.140, 141] showed, that the Shockley formula should be considered only as a rough evaluation, since at very high electric field the transport process depends on many factors including several phonon dispersion curves, phonon emission and absorption, and non-

parabolic band corrections. Good agreement between the Monte Carlo calculations and the experimental data for high-purity n-Si is seen in Fig. 21.17a. In the case of holes the saturation was neither predicted theoretically, nor found experimentally. A nearly saturated drift velocity was observed [21.122] only at room temperature, and the saturation should occur, by extrapolation, at $E \geq 2 \times 10^5 \text{ V/cm}$ with v_{sh} around 10^7 cm/s (Fig. 21.16c).

Empirically adjusted formulas for $v_d(E)$ have been proposed for varying N_i and $T = 300 \text{ K}$

$$v_d = \frac{\mu_0 E}{\sqrt{1 + \frac{S(N_i/N_{ref})}{N_i/N_{ref} + S} + \frac{(E/E_1)^2}{E/E_1 + F} + (E/E_2)^2}} \quad (21.31)$$

[21.145], and for $N_i = 0$ and varying T

$$v_d = v_m \frac{E/E_c}{[1 + (E/E_c)^\beta]^{1/\beta}} \quad (21.32)$$

[21.108]; the best-fit parameters are presented in Table 21.7. Also a formula was suggested [21.14] for the electron saturated drift velocity as a function of temperature

$$v_{se}(T) = \frac{2.4 \times 10^7 \text{ cm/sec}}{1 + 0.8 \times \exp\left(\frac{600 \text{ K}}{T}\right)} \quad (21.33)$$

The applicability of (21.32) and (21.33) is clearly demonstrated by Figs. 21.17b and 21.17c, respectively.

Hot-Carrier Diffusion

Knowledge of diffusion processes is useful for a better understanding of charge-transport phenomena and correct simulation of high-frequency devices. At low fields the diffusion coefficient D is related to the mobility by the Einstein relation (Sect. 21.1.3). At high fields it becomes a field-dependent tensor, which describes the diffusion with respect to E depending on its direction in the crystallographic frame. The difference between diffusion parallel to and perpendicular to E is greater than the variations caused by changes in the crystallographic direction. Therefore, investigations set out to determine the field dependence of both the longitudinal ($D_{||}$) and transverse (D_{\perp}) components of the diffusivity tensor with respect to E irrespective of its orientation in the crystal.

The longitudinal diffusion coefficient can be measured by the Haynes–Shockley or time-of-flight [21.120] technique by observing the difference between the fall and rise times of the current pulse. This difference is caused by the spread of the carriers traveling across the

sample and is simply related to D_{\parallel} (e.g. [21.146]). Analogously, D_{\perp} can be obtained by observing the spread of the current perpendicular to the direction of the field. The current is originated by a point excitation on one surface of a Si wafer and is collected on the opposite surface by several electrodes of appropriate geometry [21.144]; this technique is sometimes called *geometrical* [21.14]. Finally, both D_{\parallel} and D_{\perp} have been related to noise measurements, parallel and perpendicular respectively to the current direction [21.143].

Figure 21.18 shows some experimental results on the field dependence of D_{\parallel} and D_{\perp} for electrons in Si at room temperature with $\mathbf{E} \parallel \langle 111 \rangle$. The data obtained by the noise measurements are in a reasonable agreement with the time-of-flight results, although the former cover a narrower range of \mathbf{E} , just outside Ohm's region. As \mathbf{E} increases D_{\parallel} decreases to about one third of its low-field value ($\approx 36 \text{ cm}^2/\text{s}$), which is in substantial agreement with theoretical Monte Carlo computations for the nonparabolic band [21.113]. The results for transverse diffusion showed that, as \mathbf{E} increases, D_{\perp} also decreases, but to a lesser extent than D_{\parallel} . There exists a hypothesis of validity to extrapolate the Einstein relation outside the linear region by $D_E = \frac{2E}{3q} \mu(E)$. As seen from Fig. 21.18, this yields a qualitative interpretation of D_{\perp} for not too high fields. As far as D_{\parallel} is concerned, the diffusion process seems much more complex than pictured by the Einstein relation. For holes, the dependence of D_{\parallel} on \mathbf{E} was found to be similar to that for electrons [21.14].

Impact Ionization

Impact ionization is an important charge-generation mechanism. It occurs in many silicon-based devices, either determining the useful characteristic of the device or causing an unwanted parasitic effect. The breakdown of a silicon p–n diode is caused by impact ionization if its breakdown voltage is larger than about 8 V. The operation of such devices as thyristors, impact avalanche transit time (IMPATT) diodes and trapped plasma avalanche-triggered transit (TRAPATT) diodes is based on avalanche generation, the phenomenon that results from impact ionization. The avalanche generation also plays an increasing role in degradation due to hot-carrier effects and bipolar parasitic breakdown of metal–oxide–semiconductor (MOS) devices, the geometrical dimensions of which have been scaled down recently.

The ionization rate is defined as the number of electron–hole pairs generated by a carrier per unit distance traveled in a high electric field, and is different for

electrons and for holes. Impact ionization can only occur when the particle gains at least the threshold energy for ionization from the electrical field. This can be derived from the application of the energy and momentum conservation laws to the amount $E_i \approx 1.5E_g$ (assuming that the effective masses of electron and hole are equal). A large spread of experimental values for E_i exists, with a breakdown field of order of $3 \times 10^5 \text{ V/cm}$. For more detailed consideration we refer the reader to the review article [21.147] noted in Sect. 21.2.3.

21.2.3 Review Material

The following materials may be recommended for further reading.

1. Electrical properties of Si [21.24]. Summary of papers on the subject that were published over a decade until 1965 are overviewed. Miscellaneous properties, such as piezoresistance and high-electric-field mobility, were also presented.
2. Electron mobility and resistivity in n-Si versus dopant density and temperature [21.148]. An improved model for computing μ_e as a function of N_d and T in uncompensated n-Si was formulated. The effects of electron–electron interaction on conventional scattering processes, as well as their anisotropies were incorporated empirically. The model was verified to $\pm 5\%$ of the mobility measured on wafers doped by phosphorous in the range $10^{13} - 10^{19} \text{ cm}^{-3}$.
3. Bulk charge-transport properties of Si [21.14]. Review of knowledge on the subject with special emphasis on application to solid-state devices. Most attention was devoted to experimental findings at room temperatures and to high-field properties. The techniques for drift-velocity measurements and the principles of Monte Carlo simulation were overviewed. Empirical expressions were given, when possible, for the most important transport quantities as functions of T , $N_{d(a)}$ and E .
4. Semi-empirical relations for the carrier mobilities [21.149]. From a review of different publications on $\mu_{e,h}$ in Si, the authors proposed an approximated calculation procedure, analogous to that of Li and Thurber, which permits a quick and accurate evaluation of $\mu_{e(h)}$ over a wide range of T , $N_{d(a)}$ and $n(p)$. The proposed relations are well adapted to device simulation since they allow short computation times.
5. Minority-carrier recombination in heavily doped silicon [21.150]. A review of understanding of

the recombination of minority carriers in heavily doped Si. A short phenomenological description of the carrier recombination process and lifetime was provided and the main theories of these were briefly reviewed with indications for their expected contributions in heavily doped Si. The various methods used for measuring the minority-carrier lifetime in heavily doped Si were described and critically examined. The insufficiency of existing theories to explain the patterns of lifetime versus doping was clearly demonstrated.

6. Minority-carrier transport modeling in heavily doped silicon emitters [21.111]. The experimental and theoretical efforts that addressed such important issues as: (i) the incomplete understanding of the minority-carrier physics in heavily doped Si, (ii) the lack of precise measurements for the minority-carrier parameters, (iii) the difficulties encountered with the modeling of transport and recombination in nonhomogeneously doped regions, and (iv) prob-

lems with the characterization of real emitters in bipolar devices, were reviewed with the goal of being able to achieve accurate modeling of the current injected into an arbitrarily heavily doped region in a silicon device.

7. Impact ionization in silicon: a review and update [21.147]. The multiplication factor and the ionization rate were revisited. The interrelationship between these parameters together with the multiplication and breakdown models for diodes and MOS transistors were discussed. Different models were compared and test structures were discussed to measure the multiplication factor accurately enough for reliable extraction of the ionization rates. Multiplication measurements at different T were performed on a BJT, and yielded new electron ionization rates at relatively low electric fields. An explanation for the spread of existed experimental data on ionization rate was given. A new implementation method for a local avalanche model into a device simulator was presented.

21.3 Optical Properties

21.3.1 Diversity of Silicon as an Optical Material

In *dielectric*-like material $n > k$ ($\varepsilon_1 > 0$), where for transparency and opacity it is necessary that $n \gg k$ ($\varepsilon_1 \gg \varepsilon_2$) and $n \approx k$ ($\varepsilon_1 \ll \varepsilon_2$), respectively (21.16). In *metallic*-like material $n < k$ ($\varepsilon_1 < 0$), where for good reflectivity and bad reflectivity it is necessary that $n \ll k$ ($|\varepsilon_1| \gg \varepsilon_2$) and $n \approx k$ ($|\varepsilon_1| \ll \varepsilon_2$), respectively. Since in Si ε depends on the wavelength and carrier concentration, it may exhibit all these types of optical behavior ranging from dielectric-like to metallic-like. For example, (111) undoped Si ($n = 2.3 \times 10^{14} \text{ cm}^{-3}$) at $\lambda = 0.62 \mu\text{m}$ behaves as a transparent dielectric, as $\varepsilon_1 = 15.254$ and $\varepsilon_2 = 0.172$ at that wavelength, while it is an opaque dielectric at $\lambda = 0.295 \mu\text{m}$, where $\varepsilon_1 = 2.371$ and $\varepsilon_2 = 45.348$. For heavily doped Si ($n = 10^{20} \text{ cm}^{-3}$), $n = 1.911$ and $k = 8.63$ at $\lambda = 16.67 \mu\text{m}$, so it behaves as a good metallic reflector while at $\lambda = 2 \mu\text{m}$, where $n = 3.47$ and $k = 6.131 \times 10^{-3}$ it is a transparent dielectric.

21.3.2 Measurements of Optical Constants

Various methods are used to measure the dielectric constant of single-crystalline Si including transmission,

reflection, and ellipsometric methods. For a smooth opaque sample, the quantity of interest is the complex reflection amplitude ρ_r (at normal incidence) defined by

$$\begin{aligned}\rho_r &= -\frac{\sqrt{\varepsilon} - 1}{\sqrt{\varepsilon} + 1}, \\ |\rho_r|^2 &= \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} = R_0, \\ \phi &= \arctan \frac{2k}{n^2 + k^2 - 1} = \arg(\rho_r), \\ &\text{mod } (\pi).\end{aligned}\quad (21.34)$$

The Fresnel reflectance spectrum $R_0(\omega)$ is measured using reflectometry. The KKR analysis also applies to the causal function $\ln(\rho_r) = \frac{1}{2} \ln R_0 + i\phi$ that gives

$$\phi(\omega) = \frac{1}{2\pi} \int_0^\infty \ln \left| \frac{\omega + \Omega}{\omega - \Omega} \right| \frac{d}{d\Omega} \ln R_0(\Omega) d\Omega. \quad (21.35)$$

With the $\psi(\omega)$ retrieved in this way, the last two relations in (21.34) are simultaneously solved to yield the n and k spectra. The KKR method requires, in principle, data over an infinite ω range, which are supplied by measurements over a confined range and an appeal to simple models for high and low frequencies. This limi-

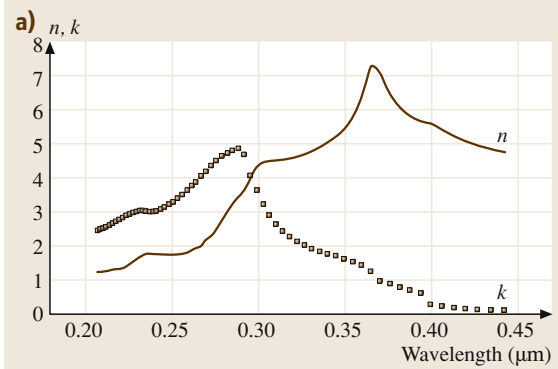


Fig. 21.19 Refractive index n and the extinction coefficient k versus wavelength, using Adachi and Geist models in the wavelength ranges: (a) 0.2–0.45 μm ,

tation is avoided using optical measurements on smooth slabs.

For a slab of thickness d , the reflectance and transmittance at normal incidence are given by

$$R = R_0 \frac{1 - 2T_0 \cos 2\varphi + T_0^2}{1 - 2R_0 T_0 \cos 2\chi + R_0^2 T_0^2},$$

$$T = T_0 \frac{1 - 2R_0 \cos 2\psi + R_0^2}{1 - 2R_0 T_0 \cos 2\chi + R_0^2 T_0^2},$$

$$T_0 = e^{-\alpha d}, \quad (21.36)$$

where $\varphi = 2\pi nd/\lambda$ is the optical-path phase, $\chi = \varphi + \psi$ and $\alpha = 4\pi kd/\lambda$ is the power attenuation coefficient. The parameter αd characterizes the slab's opacity. At $n > k$ ($\varepsilon_1 > 0$, (21.17)), i.e. in the optically dielectric range, α is referred to as *absorption coefficient*. In the optically metallic range that exists for doped Si in the IR, where $k > n$ ($\varepsilon_1 < 0$, (21.17)), α^{-1} is the skin depth multiplied by a factor of ≈ 1 . If the dispersion of the optical constants is disregarded, R and T are periodic in $1/\lambda$ (21.20). For nonopaque slabs, these interference effects show up at $nd/\lambda > 1$. For thick slabs ($nd/\lambda \gg 1$), even small fluctuation of n or/and λ causes changes in φ of the order 2π . Therefore, if such factors as bulk defects, source incoherence and low spectral resolution are present in an experiment, then the averages of R and T over φ , given by

$$\langle R \rangle = R_0 (1 + T_0 \langle T \rangle),$$

$$\langle T \rangle = T_0 \frac{1 - 2R_0 \cos 2\psi + R_0^2}{1 - R_0^2 T_0^2}, \quad (21.37)$$

are more appropriate than (21.36). This *multiple-reflection approximation* [21.151] has proved effective

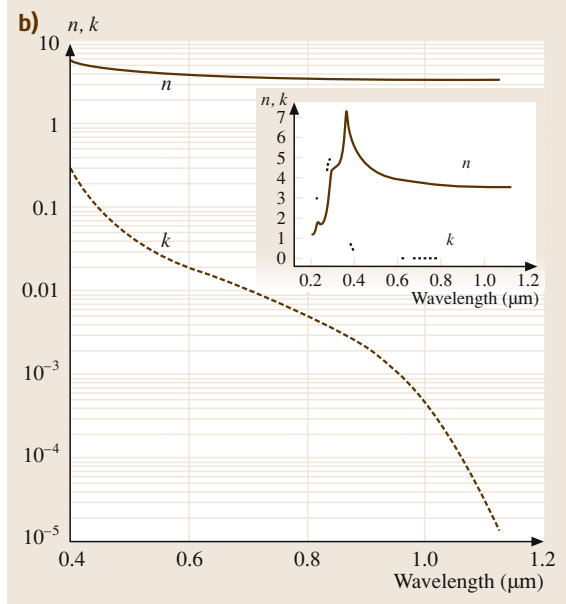


Fig. 21.19 (b) 0.4–1.127 μm . *Inset* – the complete range from 0.2 to 1.127 μm

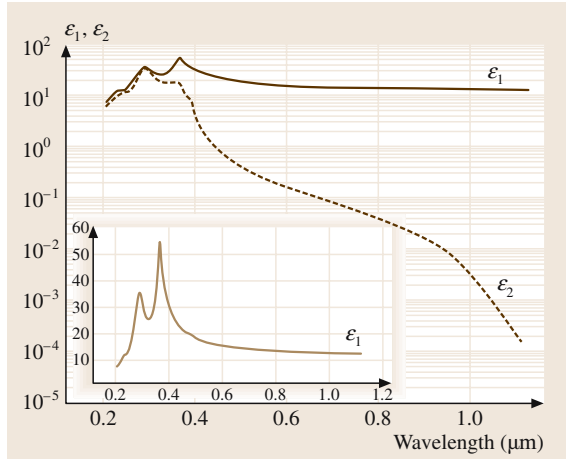


Fig. 21.20 Real ε_1 and imaginary ε_2 parts of dielectric constant versus wavelength, using Adachi and Geist models. *Inset*: real dielectric constant ε_1 versus wavelength

in countless studies. The apparent absorbance of the slab in the multiple-reflection approximation is calculated by

$$\langle A \rangle = 1 - \langle R \rangle - \langle T \rangle$$

$$= \frac{(1 - R_0)(1 - T_0) - 4R_0 T_0 \sin^2 \psi}{1 - R_0 T_0}. \quad (21.38)$$

Table 21.15 The refractive index n and the extinction coefficient k of n-Si with electron concentration $N = 10^{16} \text{ cm}^{-3}$ at various wavelength

Energy (eV)	Wavenumber (cm^{-1})	Wavelength (μm)	n HW ^a	GDA ^b	k HW ^a	GDA ^b
0.6199	5000	2.000	3.453	3.453	1.160×10^{-7}	1.514×10^{-7}
0.5579	4500	2.222	3.447	3.447	1.594×10^{-7}	2.006×10^{-7}
0.4959	4000	2.500	3.441	3.441	2.273×10^{-7}	2.751×10^{-7}
0.4339	3500	2.857	3.435	3.435	3.398×10^{-7}	3.945×10^{-7}
0.3720	3000	3.333	3.431	3.431	5.403×10^{-7}	6.001×10^{-7}
0.3100	2500	4.000	3.427	3.427	9.347×10^{-7}	9.912×10^{-7}
0.2480	2000	5.000	3.424	3.424	1.827×10^{-6}	1.850×10^{-6}
0.1860	1500	6.667	3.421	3.421	4.334×10^{-6}	4.210×10^{-6}
0.1240	1000	10.00	3.419	3.419	1.463×10^{-5}	1.392×10^{-5}
0.1116	900	11.11	3.419	3.419	2.006×10^{-5}	1.910×10^{-5}
0.09919	800	12.50	3.419	3.419	2.856×10^{-5}	2.727×10^{-5}
0.08679	700	14.29	3.418	3.418	4.262×10^{-5}	4.092×10^{-5}
0.07439	600	16.67	3.418	3.418	6.765×10^{-5}	6.549×10^{-5}
0.06199	500	20.00	3.417	3.417	1.168×10^{-4}	1.143×10^{-4}
0.04959	400	25.00	3.416	3.416	2.278×10^{-4}	2.259×10^{-4}
0.03720	300	33.33	3.413	3.413	5.382×10^{-4}	5.413×10^{-4}
0.03472	280	35.71	3.412	3.412	6.612×10^{-4}	6.669×10^{-4}
0.03224	260	38.46	3.411	3.411	8.247×10^{-4}	8.341×10^{-4}
0.02976	240	41.67	3.410	3.410	1.047×10^{-3}	1.061×10^{-3}
0.02728	220	45.45	3.408	3.408	1.356×10^{-3}	1.379×10^{-3}
0.02480	200	50.00	3.406	3.406	1.799×10^{-3}	1.834×10^{-3}
0.02232	180	55.56	3.399	3.399	3.481×10^{-3}	3.563×10^{-3}
0.01984	160	62.50	3.403	3.403	2.458×10^{-3}	2.511×10^{-3}
0.01736	140	71.43	3.394	3.394	5.154×10^{-3}	5.284×10^{-3}
0.01488	120	83.33	3.385	3.385	8.085×10^{-3}	8.297×10^{-3}
0.01240	100	100.0	3.372	3.372	1.370×10^{-2}	1.405×10^{-2}

^a Values calculated using an empirical fit [21.78], ^b values calculated using GDA [21.79]

Equation (21.38) at $k \ll n$ is extensively used in silicon-wafer thermometry [21.152, 153]. Given $\langle R \rangle$, $\langle T \rangle$ and d , (21.37) builds up a system of two equations for the two unknowns n and k . Thus, measurement of the reflection and transmission on the same slab of known thickness allows one to retrieve the optical constants. This R – T measurement method [21.154] is greatly simplified at $2k \ll n^2 + k^2 - 1$. Under this low-loss condition the above system can be solved analytically for T_0 and R_0 . The calculated T_0 directly yields k , and n is then found using the calculated R_0 . The R – T technique is the best method at $ad \leq 1$, while at $ad \gg 1$, where solving the aforementioned system becomes an ill-conditioned problem, the KKR analysis is more reliable. In two last decades, spectroscopic ellipsometry has gained wide recognition for being more precise than photometric methods. In ellipsometry, the ratio of reflectance for s - and p -polarized radiation, and the relative phase shift between

the two, are both measured at large angles of incidence [21.155].

The measured results are affected by the structural atomic-scale properties of the samples. These properties are defined by polishing processes – mechanical or chemical – that affect the surface damage and roughness, the properties of the surface native oxide, the growth mechanism of the measured layer, grain boundaries, and the quality of the cleaved surface. Since Si samples may be optically inhomogeneous, retrieving the optical constants from measurements may become a complicated inverse electromagnetic problem [21.156, 157], which is why some of the reported data for ε disagree by up to 30%. A detailed list of publications on the subject can be found in [21.158, 159]. Emphasis on these effects should be especially considered when transmission measurement is done for a wavelength range in which the absorption coefficient is large and thin samples are therefore required.

Table 21.16 The refractive index n and the extinction coefficient k of n-Si with electron concentration $N = 10^{20} \text{ cm}^{-3}$ at various wavelength

Energy (eV)	Wavenumber (cm^{-1})	Wavelength (μm)	n		k	
			HW ^a	GDA ^b	HW ^a	GDA ^b
0.6199	5000	2.000	3.270 ^P		1.834×10^{-2}	
			3.257 ^{As}	3.247	2.403×10^{-2}	6.131×10^{-3}
0.5579	4500	2.222	3.219		2.549×10^{-2}	
			3.203	3.190	3.341×10^{-2}	9.701×10^{-3}
0.4959	4000	2.500	3.151		3.698×10^{-2}	
			3.130	3.112	4.845×10^{-2}	1.664×10^{-2}
0.4339	3500	2.857	3.053		5.671×10^{-2}	
			3.027	3.000	7.432×10^{-2}	3.232×10^{-2}
0.3720	3000	3.333	2.902		9.410×10^{-2}	
			2.867	2.828	1.233×10^{-1}	8.053×10^{-2}
0.3100	2500	4.000	2.644		1.765×10^{-1}	
			2.597	2.577	2.314×10^{-1}	2.549×10^{-1}
0.2480	2000	5.000	2.138		4.176×10^{-1}	
			2.087	1.990	5.461×10^{-1}	5.336×10^{-1}
0.2356	1900	5.263	1.981		5.226×10^{-1}	
			1.939		6.796×10^{-1}	
0.2232	1800	5.556	1.800		6.717×10^{-1}	
			1.780		8.617×10^{-1}	
0.2108	1700	5.882	1.604		8.873×10^{-1}	
			1.626		1.107×10^0	
0.1984	1600	6.250	1.423		1.188×10^0	
			1.503		1.417×10^0	
0.1860	1500	6.667	1.295		1.566×10^0	
			1.429	0.971	1.777×10^0	1.902×10^0
0.1736	1400	7.143	1.237		1.989×10^0	
			1.411	0.893	2.172×10^0	2.369×10^0
0.1612	1300	7.692	1.239		2.437×10^0	
			1.442	0.873	2.591×10^0	2.859×10^0
0.1488	1200	8.333	1.291		2.911×10^0	
			1.518	0.872	3.037×10^0	3.387×10^0
0.1364	1100	9.091	1.390		3.418×10^0	
			1.642	0.913	3.514×10^0	3.960×10^0
0.1240	1000	10.00	1.540		3.971×10^0	
			1.817	0.985	4.034×10^0	4.603×10^0

^a Values calculated using an empirical fit [21.78], ^b values calculated using GDA [21.79]

21.3.3 Modeling of Optical Constants

A method for calculating ε_1 and ε_2 and then n , k and α for photon energies of 0–6 eV ($\lambda > 0.2 \mu\text{m}$) has been reported [21.82]. The calculated data are in excellent agreement with experimental data for the wavelength range 0.2–4 μm [21.156, 160]. The model is based on the KKR (Sect. 21.1.3) and takes into

account the dependence of ε on the energy-band structure. It considers the effect of indirect-band-gap and inter-band transitions as well as the electron (conduction bands) and hole (valence bands) density of states. The fundamental absorption (generation of electron–hole pair) edge energy of 1.12 eV corresponds to the indirect transition from the highest valence band to the lowest conduction band. Sharp changes in the

Table 21.17 Table 21.16 cont.

Energy (eV)	Wavenumber (cm ⁻¹)	Wavelength (μm)	<i>n</i> HW ^a	GDA ^b	<i>k</i> HW ^a	GDA ^b
0.1116	900	11.11	1.751 ^P 2.057 ^{As}	1.105	4.584 × 10 ⁰ 4.608 × 10 ⁰	5.341 × 10 ⁰
0.09919	800	12.50	2.042 2.378	1.278	5.278 × 10 ⁰ 5.251 × 10 ⁰	6.216 × 10 ⁰
0.08679	700	14.29	2.444 2.810	1.530	6.079 × 10 ⁰ 5.982 × 10 ⁰	7.282 × 10 ⁰
0.08431	680	14.71	2.542 2.913		6.255 × 10 ⁰ 6.141 × 10 ⁰	
0.08183	660	15.15	2.646 3.022		6.437 × 10 ⁰ 6.305 × 10 ⁰	
0.07935	640	15.63	2.758 3.139		6.625 × 10 ⁰ 6.474 × 10 ⁰	
0.07690	620	16.13	2.877 3.262		6.820 × 10 ⁰ 6.649 × 10 ⁰	
0.07439	600	16.67	3.006 3.395	1.911	7.022 × 10 ⁰ 6.829 × 10 ⁰	8.630 × 10 ⁰
0.07180	580	17.24	3.143 3.536		7.231 × 10 ⁰ 7.015 × 10 ⁰	
0.06933	560	17.86	3.291 3.686		7.448 × 10 ⁰ 7.208 × 10 ⁰	
0.06685	540	18.52	3.451 3.846		7.674 × 10 ⁰ 7.407 × 10 ⁰	
0.06438	520	19.23	3.622 4.018		7.909 × 10 ⁰ 7.614 × 10 ⁰	
0.06199	500	20.00	3.807 4.201	2.517	8.154 × 10 ⁰ 7.829 × 10 ⁰	1.040 × 10 ¹
0.05951	480	20.83	4.006 4.398		8.409 × 10 ⁰ 8.052 × 10 ⁰	
0.05703	460	21.74	4.222 4.608		8.675 × 10 ⁰ 8.285 × 10 ⁰	
0.05455	440	22.73	4.455 4.835		8.954 × 10 ⁰ 8.528 × 10 ⁰	
0.05207	420	23.81	4.707 5.078		9.245 × 10 ⁰ 8.782 × 10 ⁰	
0.04959	400	25.00	4.982 5.340	3.551	9.551 × 10 ⁰ 9.048 × 10 ⁰	1.285 × 10 ¹

^a Values calculated using an empirical fit [21.78], ^b values calculated using GDA [21.79]

optical constants are obtained at wavelengths around 0.367, 0.29 and 0.233 μm, which correspond to the energy-band critical points of 3.38, 4.27 and 5.317 eV, respectively. An additional analytical model for calculating the *n* and *k* values for the wavelength range 0.4–1.127 μm has been developed [21.158]. The model

is based on measured *k* and *n* data [21.161–163], where the calculated values are within ±10% of the measured values. Using Adachi and Geist models we have calculated and plotted *n*, *k*, ε₁ and ε₂ for the wavelength range 0.2–1.127 μm, as seen in Figs. 21.19 and 21.20.

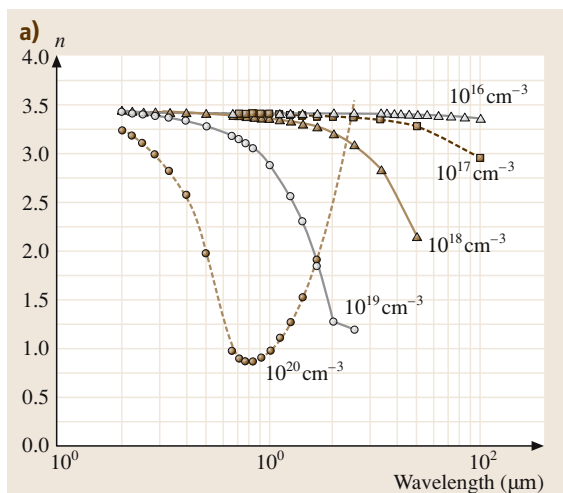


Fig. 21.21 Optical constants versus wavelength at various electron concentrations using the GDA model. (a) Refractive index n .

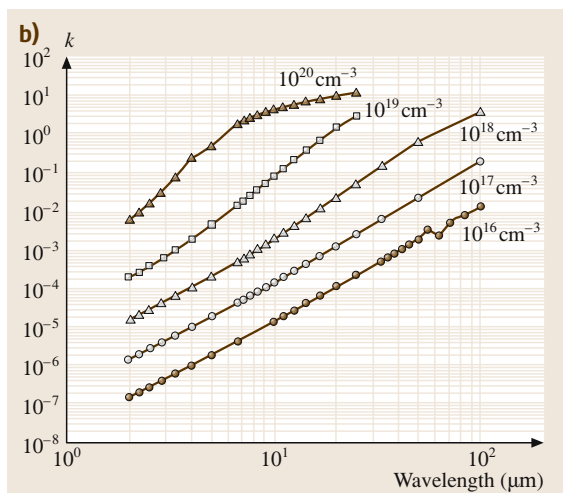


Fig. 21.21 (b) Extinction coefficient k

In the IR range the lattice and free carriers, which arise from doping, give additive contributions. The free-carrier contribution to the dielectric function when the radiation quantum is much smaller than the mean electron energy, $\hbar\omega \ll \bar{E}$, is considered using the semiclassical Drude approximation (DA, Sect. 21.1.4), which leads to the simple formula

$$\varepsilon = \varepsilon_L - \frac{n_L^2 \omega_p^2}{\omega(\omega + i\gamma)}, \quad \gamma = \frac{1}{\tau}, \quad (21.39)$$

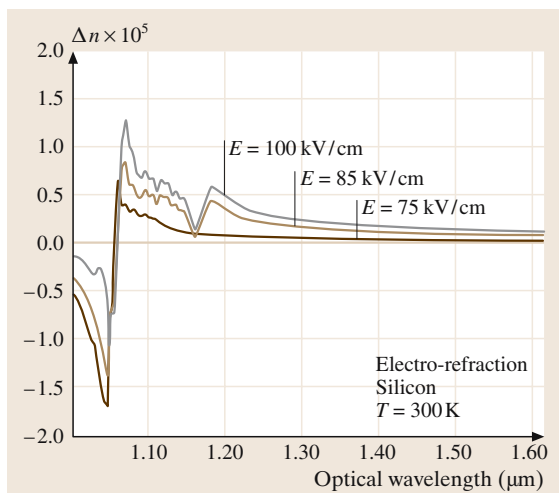


Fig. 21.22 Refractive index change versus wavelength at various electric fields [21.164]

where τ is the phenomenological relaxation time, ω_p is the electron plasma frequency given by

$$\omega_p = \sqrt{\frac{4\pi q^2 N}{\varepsilon_L m_c}}. \quad (21.40)$$

Here N denotes the carriers' concentration, $m_c \approx 0.26m_0$ for electrons (Table 21.2) and $m_c = (0.33-0.39)m_0$ for holes (Sect. 21.1.4).

Advanced GDA, which extrapolates the free-carrier contribution to the range of $\hbar\omega > \bar{E}$ was considered [21.79]. In Table 21.15 we tabulated the values of n and k for practical use at two extreme electron concentrations: 10^{16} cm^{-3} and 10^{20} cm^{-3} . The results placed in the columns headed by the abbreviation 'HW' were obtained using DA (21.21, 40, 41) and the empirically adjusted τ versus N fit [21.78]. The columns of theoretical values calculated by us using GDA are headed by the corresponding abbreviation. A disagreement between the theory and the empirical fit notable at heavy doping has been discussed in detail [21.79]. The dependencies of n and k on wavelength in the range at carrier concentration of $10^{16}-10^{20} \text{ cm}^{-3}$ in steps of 10 cm^{-3} are shown in Fig. 21.21.

21.3.4 Electric-Field and Temperature Effects on Optical Constants

It has been known for many years that the optical absorption spectrum of Si can also be affected by external electric fields (the Franz-Keldysh effect). The Franz-

Keldysh effect, which alters the α spectrum of crystalline Si, is field-induced tunneling between valence- and conduction-band states. In recent years, the generic term *electroabsorption* has been adopted for $\Delta\alpha$ versus E effects. The effect of electric field on the refractive index is shown in Fig. 21.22 [21.164]. Sharp changes occur around the wavelengths correspond to the band-gap transition.

The temperature dependence of the refractive index of high-purity damage-free Si, for photon energies less than 3 eV in the temperature range 300–500 K is given by [21.162]

$$\frac{\Delta n}{\Delta T} \approx 1.3 \times 10^{-4} n \quad (\text{K}^{-1}). \quad (21.41)$$

References

- 21.1 C. Kittel: *Introduction to Solid State Physics*, 6th edn. (Wiley, New York 1986)
- 21.2 C. Kittel: *Quantum Theory of Solids*, 2nd edn. (Wiley, New York 1987)
- 21.3 K. Seeger: *Semiconductor Physics* (Springer, New York 1982)
- 21.4 T. S. Moss: *Optical Properties of Semiconductors* (Butterworths, London 1959)
- 21.5 S. M. Sze: *Physics of Semiconductor Devices* (Wiley, New York 1981)
- 21.6 H. M. van Driel: *Appl. Phys. Lett.* **44**, 617 (1984)
- 21.7 E. O. Kane: *J. Phys. Chem. Solids* **1**, 82 (1956)
- 21.8 J. R. Chelikowsky, M. L. Cohen: *Phys. Rev. B* **14**, 556 (1976)
- 21.9 B. Lax, J. G. Mavroides: *Phys. Rev.* **100**, 1650 (1955)
- 21.10 J. C. Hensel, G. Feher: *Phys. Rev.* **129**, 1041 (1963)
- 21.11 I. Balslev, P. Lawaetz: *Phys. Lett.* **19**, 3460 (1965)
- 21.12 P. Lawaetz: *Phys. Rev. B* **4**, 3460 (1971)
- 21.13 G. Ottaviani, L. Reggiani, C. Canali, F. Nava, A. A-Quranta: *Phys. Rev. B* **12**, 3318 (1975)
- 21.14 C. Jacoboni, C. Canali, G. Ottaviani, A. A-Quranta: *Solid State Electron.* **20**, 77 (1977)
- 21.15 H. D. Barber: *Solid State Electron.* **10**, 1039 (1967)
- 21.16 M. A. Green: *J. Appl. Phys.* **67**, 2944 (1990)
- 21.17 A. B. Sproul, M. A. Green: *J. Appl. Phys.* **70**, 846 (1991)
- 21.18 R. F. Pierret: *Advanced Semiconductor Fundamentals*, Modular Series on Solid State Devices, ed. by G. W. Neudeck, R. F. Pierret (Pearson Education, New York 2003)
- 21.19 G. Dresselhaus, A. F. Kip, C. Kittel: *Phys. Rev.* **98**, 368 (1955)
- 21.20 J. C. Hensel, H. Hasegawa, M. Nakayama: *Phys. Rev.* **138**, 225 (1965)
- 21.21 L. Kleinmann, J. C. Phillips: *Phys. Rev.* **118**, 1153 (1960)
- 21.22 M. Cardona, F. H. Pollak: *Phys. Rev.* **142**, 530 (1966)
- 21.23 S. Zwerdling, K. J. Button, B. Lax, L. M. Roth: *Phys. Rev. Lett.* **4**, 173 (1960)
- 21.24 W. R. Runyan: *Silicon Semiconductor Technology* (McGraw-Hill, New York 1965) Chap. 8
- 21.25 V. I. Fistul: *Heavily Doped Semiconductors* (Plenum, New York 1969)
- 21.26 W. Kohn, J. M. Luttinger: *Phys. Rev.* **97**, 1721 (1955)
- 21.27 W. Kohn, J. M. Luttinger: *Phys. Rev.* **98**, 915 (1955)
- 21.28 W. Kohn, D. Schechter: *Phys. Rev.* **99**, 1903 (1955)
- 21.29 E. Burstein, G. Picus, B. Henvi, R. Wallis: *J. Phys. Chem. Solids* **1**, 65 (1956)
- 21.30 G. Picus, E. Burstein, B. Henvi: *J. Phys. Chem. Solids* **1**, 75 (1956)
- 21.31 N. F. Mott: *Metal-Insulator Transitions*, 2nd edn. (Taylor & Francis, London 1990) p. 2
- 21.32 R. H. Hall, J. H. Racette: *J. Appl. Phys.* **35**, 379 (1964)
- 21.33 W. M. Bullis: *Solid State Electron.* **9**, 143 (1966)
- 21.34 B. Jensen: *Handbook of Optical Constants of Solids*, Vol. 2 (Academic, Orlando 1985) p. 169
- 21.35 M. Kohler: *Z. Physik* **124**, 777 (1948)
- 21.36 M. Kohler: *Z. Physik* **125**, 679 (1949)
- 21.37 B. R. Nag: *Theory of Electrical Transport in Semiconductors* (Pergamon, Oxford 1972)
- 21.38 E. M. Conwell: *High Field Transport in Semiconductors* (Academic, New York 1967)
- 21.39 W. Fawcett, A. D. Boardman, S. Swain: *J. Phys. Chem. Solids* **31**, 1963 (1970)
- 21.40 W. Dumke: *Phys. Rev.* **124**, 1813 (1961)
- 21.41 I. Balslev: *Phys. Rev.* **143**, 636 (1966)
- 21.42 B. N. Brockhouse: *Phys. Rev. Lett.* **2**, 256 (1959)
- 21.43 W. A. Harrison: *Phys. Rev.* **104**, 1281 (1956)
- 21.44 H. Ehrenreich, A. W. Overhauser: *Phys. Rev.* **104**, 331 (1956)
- 21.45 J. Bardeen, W. Shockley: *Phys. Rev.* **80**, 72 (1950)
- 21.46 G. L. Bir, G. E. Pikus: *Fiz. Tverd. Tela* **22**, 2039 (1960) *Soviet Phys. - Solid State* **2** (1961) 2039
- 21.47 M. Tiersten: *IBM J. Res. Devel.* **5**, 122 (1961)
- 21.48 C. Herring, E. Vogt: *Phys. Rev.* **101**, 944 (1956)
- 21.49 M. Tiersten: *J. Phys. Chem. Solids* **25**, 1151 (1964)
- 21.50 H. W. Streitwolf: *Phys. Stat. Sol.* **37**, K47 (1970)
- 21.51 D. Long: *Phys. Rev.* **120**, 2024 (1960)
- 21.52 D. L. Rode: *Phys. Stat. Sol. (b)* **53**, 245 (1972)
- 21.53 P. Norton, T. Braggins, H. Levinstein: *Phys. Rev. B* **8**, 5632 (1973)
- 21.54 C. Canali, C. Jacobini, F. Nava, G. Ottaviani, A. Alberigi: *Phys. Rev. B* **12**, 2265 (1975)
- 21.55 C. Erginsoy: *Phys. Rev.* **79**, 1013 (1950)
- 21.56 N. Sclar: *Phys. Rev.* **104**, 1559 (1956)
- 21.57 E. M. Conwell, V. F. Weisskopf: *Phys. Rev.* **77**, 338 (1950)
- 21.58 H. Brooks: *Phys. Rev.* **83**, 388 (1951)
- 21.59 C. Herring: *Bell Syst. Tech. J.* **36**, 237 (1955)
- 21.60 R. Dingle: *Phil. Mag.* **46**, 831 (1955)

- 21.61 H. Brooks: *Advances in Electronics and Electron Physics*, Vol. 7 (Academic, New York 1955) p. 85
- 21.62 D. Chattopadhyay, H. J. Queisser: *Rev. Mod. Phys.* **53**, 745 (1981)
- 21.63 A. G. Samoilovich, I. Ya. Korenblit, I. V. Dakhovskii, V. D. Iskra: *Fiz. Tverd. Tela* **3**, 3285 (1961) *Soviet Phys. – Solid State* (1962) 2385
- 21.64 P. M. Eagles, D. M. Edwards: *Phys. Rev.* **138**, A1706 (1965)
- 21.65 I. V. Dakhovskii: *Fiz. Tverd. Tela* **55**, 2332 (1963) *Soviet Phys. – Solid State* **5** (1964) 1695
- 21.66 D. Long, J. Myers: *Phys. Rev.* **120**, 39 (1960)
- 21.67 J. R. Meyer, F. J. Bartoli: *Phys. Rev. B* **23**, 5413 (1981)
- 21.68 H. I. Ralph, G. Simpson, R. J. Elliot: *Phys. Rev. B* **11**, 2948 (1975)
- 21.69 H. K. Jung, H. Ohtsuka, K. Taniguchi, C. Hamaguchi: *J. Appl. Phys.* **79**, 2559 (1996)
- 21.70 G. L. Bir, E. Normantas, G. E. Pikus: *Fiz. Tverd. Tela* **4**, 1180 (1962)
- 21.71 P. P. Debye, E. M. Conwell: *Phys. Rev.* **93**, 693 (1954)
- 21.72 J. Appel: *Phys. Rev.* **122**, 1760 (1961)
- 21.73 M. Luong, A. W. Shaw: *Phys. Rev. B* **4**, 30 (1971)
- 21.74 J. Appel: *Phys. Rev.* **125**, 1815 (1962)
- 21.75 B. E. Sernelius: *Phys. Rev. B* **41**, 2436 (1990)
- 21.76 P. A. Shumann, R. P. Phillips: *Solid State Electron.* **10**, 943 (1967)
- 21.77 M. A. Saifi, R. H. Stolen: *J. Appl. Phys.* **43**, 1171 (1972)
- 21.78 J. Humlíček, K. Wojtechovsky: *Czech. J. Phys. B* **38**, 1033 (1988)
- 21.79 M. Auslender, S. Hava: *Handbook of Optical Constants of Solids*, Vol. 3, ed. by D. Palik E. (Academic, New York 1998) p. 155
- 21.80 W. Kaiser, P. H. Keck, C. F. Lange: *Phys. Rev.* **101**, 1264 (1956)
- 21.81 S. Adachi: *Phys. Rev. B* **38**, 12966 (1988)
- 21.82 S. Adachi: *J. Appl. Phys.* **66**, 3224 (1989)
- 21.83 T. Aoki, S. Adachi: *J. Appl. Phys.* **69**, 1574 (1991)
- 21.84 K.-F. Berggren, B. E. Sernelius: *Phys. Rev. B* **24**, 1971 (1981)
- 21.85 P. E. Schmid: *Phys. Rev. B* **23**, 5531 (1981)
- 21.86 L. Viña, M. Cardona: *Phys. Rev. B* **29**, 6739 (1984)
- 21.87 W. G. Spitzer, H. Y. Fan: *Phys. Rev.* **106**, 882 (1957)
- 21.88 M. Balkanski, A. Aziza, E. Amzallag: *Phys. Stat. Sol.* **31**, 323 (1969)
- 21.89 M. Cardona, W. Paul, H. Brooks: *Zeitschr. Naturforsch.* **101**, 329 (1960)
- 21.90 L. M. Lambert: *Phys. Stat. Sol.* **11**, 461 (1972)
- 21.91 L. Pearson, J. Bardeen: *Phys. Rev.* **75**, 865 (1961)
- 21.92 M. B. Prince: *Phys. Rev.* **93**, 1204 (1954)
- 21.93 G. W. Ludwig, R. L. Watters: *Phys. Rev.* **101**, 1699 (1956)
- 21.94 D. C. Cronemeyer: *Phys. Rev.* **105**, 522 (1957)
- 21.95 P. P. Debye, T. Kohane: *Phys. Rev.* **94**, 724 (1954)
- 21.96 F. J. Morin, J. P. Maita: *Phys. Rev.* **96**, 28 (1954)
- 21.97 G. Backenstoss: *Phys. Rev.* **108**, 579 (1957)
- 21.98 D. Long: *Phys. Rev.* **107**, 672 (1957)
- 21.99 D. Long, J. Myers: *Phys. Rev.* **109**, 1098 (1958)
- 21.100 R. A. Logan, A. J. Peters: *J. Appl. Phys.* **31**, 122 (1960)
- 21.101 D. Long, J. Myers: *Phys. Rev.* **115**, 1107 (1959)
- 21.102 L. J. Neuringer, D. Long: *Phys. Rev.* **135**, A788 (1964)
- 21.103 A. G. Samoilovich, I. Ya. Korenblit, I. V. Dakhovskii, V. D. Iskra: *Fiz. Tverd. Tela* **3**, 2939 (1961) *Soviet Phys. – Solid State* **3** (1962) 2148
- 21.104 J. C. Irvin: *Bell Syst. Tech. J.* **41**, 387 (1962)
- 21.105 F. Mousty, P. Ostojá, L. Passari: *J. Appl. Phys.* **45**, 4576 (1974)
- 21.106 I. G. Kirnas, P. M. Kurilo, P. G. Litovchenko, V. S. Lutsyak, V. M. Nitsovich: *Phys. Stat. Sol. (a)* **23**, K123 (1974)
- 21.107 S. M. Sze, J. C. Irvin: *Solid State Electron.* **11**, 559 (1968)
- 21.108 D. M. Caughey, R. F. Thomas: *Proc. IEEE* **55**, 2192 (1967)
- 21.109 G. Baccarani, P. Ostojá: *Solid State Electron.* **18**, 1039 (1975)
- 21.110 C. Hilsum: *Electron. Lett.* **10**, 259 (1974)
- 21.111 J. A. del Alamo, R. M. Swanson: *Solid State Electron.* **30**, 1127 (1987)
- 21.112 Y. Furukawa: *J. Phys. Soc. Japan* **16**, 577 (1961)
- 21.113 C. Canali, C. Jacoboni, G. Ottaviani, A. Alberigi Quaranta: *Appl. Phys. Lett.* **27**, 278 (1975)
- 21.114 C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, A. Alberigi Quaranta: *Phys. Rev. B* **12**, 2265 (1975)
- 21.115 E. H. Putley, W. H. Mitchell: *Proc. Phys. Soc. (London) A* **72**, 193 (1958)
- 21.116 C. Canali, M. Costato, G. Ottaviani, L. Reggiani: *Phys. Rev. Lett.* **31**, 536 (1973)
- 21.117 E. A. Davies, D. S. Gosling: *J. Phys. Chem. Solids* **23**, 413 (1962)
- 21.118 C. Canali, G. Ottaviani, A. Alberigi Quaranta: *J. Phys. Chem. Solids* **32**, 1707 (1971)
- 21.119 C. B. Norris, J. F. Gibbons: *IEEE Trans. Electron. Dev.* **14**, 30 (1967)
- 21.120 T. W. Sigmon, J. F. Gibbons: *Appl. Phys. Lett.* **15**, 320 (1969)
- 21.121 V. Rodriguez, H. Ruegg, M.-A. Nicolet: *IEEE Trans. Electron. Dev.* **14**, 44 (1967)
- 21.122 T. E. Seidel, D. L. Scharfetter: *J. Phys. Chem. Solids* **28**, 2563 (1967)
- 21.123 V. Rodriguez, M.-A. Nicolet: *J. Appl. Phys.* **40**, 496 (1969)
- 21.124 B. L. Boichenko, V. M. Vasetskii: *Soviet Phys. Solid State* **7**, 1631 (1966)
- 21.125 A. C. Prior: *J. Phys. Chem. Solids* **12**, 175 (1959)
- 21.126 C. Y. Duh, J. L. Moll: *IEEE Trans. Electron. Dev.* **14**, 46 (1967)
- 21.127 C. Y. Duh, J. L. Moll: *Solid State Electron.* **11**, 917 (1968)
- 21.128 M. H. Jorgensen, N. O. Gram, N. I. Meyer: *Solid-State Comm.* **10**, 337 (1972)
- 21.129 M. Asche, B. L. Boichenko, O. G. Sarbej: *Phys. Stat. Sol.* **9**, 323 (1965)
- 21.130 J. G. Nash, J. W. Holm-Kennedy: *Appl. Phys. Lett.* **24**, 139 (1974)
- 21.131 J. G. Nash, J. W. Holm-Kennedy: *Appl. Phys. Lett.* **25**, 507 (1974)

- 21.132 M. Asche, O. G. Sarbej: Phys. Stat. Sol. (a) **38**, K61 (1971)
- 21.133 N. O. Gram: Phys. Lett. A **38**, 235 (1972)
- 21.134 C. Canali, A. Loria, F. Nava, G. Ottaviani: Solid-State Comm. **12**, 1017 (1973)
- 21.135 M. Asche, O. G. Sarbej: Phys. Stat. Sol. (a) **46**, K121 (1971)
- 21.136 J. P. Nougier, M. Rolland, O. Gasquet: Phys. Rev. B **11**, 1497 (1975)
- 21.137 E. J. Ryder, W. Shockley: Phys. Rev. **81**, 139 (1951)
- 21.138 W. Shockley: Bell. Syst. Tech. J. **30**, 990 (1951)
- 21.139 E. J. Ryder: Phys. Rev. **90**, 766 (1953)
- 21.140 M. Costato, L. Reggiani: Lett. Nuovo Cimento **3**, 728 (1970)
- 21.141 C. Jacoboni, R. Minder, G. Majni: J. Phys. Chem. Solids **36**, 1129 (1975)
- 21.142 C. Canali, G. Ottaviani: Phys. Lett. A **32**, 147 (1970)
- 21.143 J. P. Nougier, M. Rolland: Phys. Rev. B **8**, 5728 (1973)
- 21.144 G. Persky, D. J. Bartelink: J. Appl. Phys. **42**, 4414 (1971)
- 21.145 D. L. Scharfetter, H. K. Gummel: IEEE Trans. Electron. Dev. **ED-16**, 64 (1969)
- 21.146 J. G. Ruch, G. S. Kino: Phys. Rev. **174**, 921 (1968)
- 21.147 W. Maes, K. de Meyer, R. van Overstraeten: Solid State Electron. **33**, 705 (1990)
- 21.148 S. S. Li, W. R. Thurber: Solid State Electron. **20**, 609 (1977)
- 21.149 J. M. Dorkel, P. Leturcq: Solid-State Electron. **24**, 821 (1981)
- 21.150 M. S. Tyagi, R. van Overstraeten: Solid State Electron. **10**, 1039 (1983)
- 21.151 H. O. McMahon: J. Opt. Soc. Am. **40**, 376 (1950)
- 21.152 J. C. Sturm, C. M. Reaves: IEEE Trans. Electron. Dev. **39**, 81 (1992)
- 21.153 P. J. Timans: J. Appl. Phys. **74**, 6353 (1993)
- 21.154 P. A. Shumann Jr., W. A. Keenan, A. H. Tong, H. H. Gegenwarth, C. P. Schneider: J. Electrochem. Soc. **118**, 145 (1971)
- 21.155 R. M. A. Azzam, N. M. Bashara: *Ellipsometry and Polarized Light*, 2nd edn. (Elsevier, Amsterdam 1987)
- 21.156 D. E. Aspnes, A. A. Studna: Phys. Rev. B **27**, 985 (1983)
- 21.157 E. Barta, G. Lux: J. Phys. D: Appl. Phys. **16**, 1543 (1983)
- 21.158 J. Geist: *Handbook of Optical Constants of Solids*, Vol. 3, ed. by D. Palik E. (Academic, New York 1998) p. 519
- 21.159 D. E. Aspnes, A. A. Studna, E. Kinsbron: Phys. Rev. B **29**, 768 (1984)
- 21.160 H. R. Philipp, E. A. Taft: Phys. Rev. B **120**, 37 (1960)
- 21.161 G. E. Jellison Jr.: Opt. Mater. **1**, 41 (1992)
- 21.162 H. A. Weakliem, D. Redfield: J. Appl. Phys. **50**, 1491 (1979)
- 21.163 D. F. Edward: *Handbook of Optical Constants of Solids*, Vol. 2 Orlando 1985) p. 547
- 21.164 R. A. Soref, B. R. Bennett: IEEE J. Quantum Electron. **23**, 123 (1987)

22. Silicon–Germanium: Properties, Growth and Applications

Silicon–germanium is an important material that is used for the fabrication of SiGe heterojunction bipolar transistors and strained Si metal–oxide–semiconductor (MOS) transistors for advanced complementary metal–oxide–semiconductor (CMOS) and BiCMOS (bipolar CMOS) technologies. It also has interesting optical properties that are increasingly being applied in silicon-based photonic devices. The key benefit of silicon–germanium is its use in combination with silicon to produce a heterojunction. Strain is incorporated into the silicon–germanium or the silicon during growth, which also gives improved physical properties such as higher values of mobility. This chapter reviews the properties of silicon–germanium, beginning with the electronic properties and then progressing to the optical properties. The growth of silicon–germanium is considered, with particular emphasis on the chemical vapour deposition technique and selective epitaxy. Finally, the properties of polycrystalline silicon–germanium are discussed in the context of its use as a gate material for MOS transistors.

22.1	Physical Properties of Silicon–Germanium	482
22.1.1	Critical Thickness	482
22.1.2	Band Structure	483

Silicon–germanium ($\text{Si}_{1-x}\text{Ge}_x$) alloys have been researched since the late 1950s [22.1], but it is only in the past 15 years or so that these layers have been applied to new types of transistor technology. $\text{Si}_{1-x}\text{Ge}_x$ was first applied in bipolar technologies [22.2, 3], but more recently has been applied to metal–oxide–semiconductor (MOS) technologies [22.4–7]. This has been made possible by the development of new growth techniques, such as molecular-beam epitaxy (MBE), low-pressure chemical vapour deposition (LPCVD) and ultra-high-vacuum chemical vapour deposition (UHV-CVD). The key feature of these techniques that has led to the development of $\text{Si}_{1-x}\text{Ge}_x$ transistors is the growth of epitaxial layers at low temperatures (500–700 °C). This allows

22.1.3	Dielectric Constant	484
22.1.4	Density of States	484
22.1.5	Majority-Carrier Mobility in Strained $\text{Si}_{1-x}\text{Ge}_x$	486
22.1.6	Majority-Carrier Mobility in Tensile-Strained Si on Relaxed $\text{Si}_{1-x}\text{Ge}_x$	486
22.1.7	Minority-Carrier Mobility in Strained $\text{Si}_{1-x}\text{Ge}_x$	486
22.1.8	Apparent Band-Gap Narrowing in $\text{Si}_{1-x}\text{Ge}_x$ HBTs	487
22.2	Optical Properties of SiGe	488
22.2.1	Dielectric Functions and Interband Transitions	488
22.2.2	Photoluminescence	489
22.2.3	SiGe Quantum Wells	490
22.3	Growth of Silicon–Germanium	492
22.3.1	In-Situ Hydrogen Bake	492
22.3.2	Hydrogen Passivation	492
22.3.3	Ultra-Clean Epitaxy Systems	492
22.3.4	$\text{Si}_{1-x}\text{Ge}_x$ Epitaxy	492
22.3.5	Selective $\text{Si}_{1-x}\text{Ge}_x$ Epitaxy	492
22.4	Polycrystalline Silicon–Germanium	494
22.4.1	Electrical Properties of Polycrystalline $\text{Si}_{1-x}\text{Ge}_x$	496
	References	497

$\text{Si}_{1-x}\text{Ge}_x$ layers to be grown without disturbing the doping profiles of structures already present in the silicon wafer. $\text{Si}_{1-x}\text{Ge}_x$ layers can be successfully grown on silicon substrates even though there is a lattice mismatch between silicon and germanium of 4.2%.

The primary property of $\text{Si}_{1-x}\text{Ge}_x$ that is of interest for bipolar transistors is the band gap, which is smaller than that of silicon and controllable by varying the germanium content. Band-gap engineering concepts, which were previously only possible in compound semiconductor technologies, have now become viable in silicon technology. These concepts have introduced new degrees of freedom in the design of bipolar transistors that have led to dramatic improvements in transistor

performance. In $\text{Si}_{1-x}\text{Ge}_x$ heterojunction bipolar transistors (HBTs), the $\text{Si}_{1-x}\text{Ge}_x$ layer is incorporated into the base and the lower band gap of $\text{Si}_{1-x}\text{Ge}_x$ than Si is used to advantage to dramatically improve the high-frequency performance. $\text{Si}_{1-x}\text{Ge}_x$ HBTs have been produced with values of cut-off frequency, f_T , approaching 300 GHz [22.8], a value unimaginable in silicon bipolar transistors. Values of gate delay well below 10 ps can be achieved in properly optimised $\text{Si}_{1-x}\text{Ge}_x$ HBTs [22.9]. In $\text{Si}_{1-x}\text{Ge}_x$ MOS field-effect transistors (MOSFETs), the $\text{Si}_{1-x}\text{Ge}_x$ layer is incorporated in the channel and is used to give improved values of mobility.

Initially, strained $\text{Si}_{1-x}\text{Ge}_x$ layers on silicon substrates were used to give improved hole mobility in p-channel transistors [22.7], but more recently thin, strained silicon layers on relaxed SiGe virtual substrates have been used to give improvements in both electron and hole mobility [22.4–6].

In this chapter, the properties of single-crystal silicon–germanium will first be outlined, followed by a description of the methods used for growing silicon–germanium layers. The properties and applications of polycrystalline silicon–germanium are also discussed later in the article.

22.1 Physical Properties of Silicon–Germanium

Silicon and germanium are completely miscible over the full range of compositions and hence can be combined to form $\text{Si}_{1-x}\text{Ge}_x$ alloys with the germanium content, x , ranging from 0 to 1 (0–100%). $\text{Si}_{1-x}\text{Ge}_x$ has a diamond-like lattice structure and the lattice constant is given by Vegard's rule:

$$a_{\text{Si}_{1-x}\text{Ge}_x} = a_{\text{Si}} + x(a_{\text{Ge}} - a_{\text{Si}}), \quad (22.1)$$

where x is the germanium fraction and a is the lattice constant. The lattice constant of silicon, a_{Si} , is 0.543 nm, the lattice constant of germanium, a_{Ge} , is 0.566 nm and the lattice mismatch is 4.2%.

When a $\text{Si}_{1-x}\text{Ge}_x$ layer is grown on a silicon substrate, the lattice mismatch at the interface between the $\text{Si}_{1-x}\text{Ge}_x$ and the silicon has to be accommodated. This

can either be done by compression of the $\text{Si}_{1-x}\text{Ge}_x$ layer so that it fits to the silicon lattice or by the creation of misfit dislocations at the interface. These two possibilities are illustrated schematically in Fig. 22.1. In the former case, the $\text{Si}_{1-x}\text{Ge}_x$ layer adopts the silicon lattice spacing in the plane of the growth and hence the normally cubic $\text{Si}_{1-x}\text{Ge}_x$ crystal is distorted. When $\text{Si}_{1-x}\text{Ge}_x$ growth occurs in this way, the $\text{Si}_{1-x}\text{Ge}_x$ layer is under compressive strain and the layer is described as *pseudomorphic*. In the second case, the $\text{Si}_{1-x}\text{Ge}_x$ layer is unstrained, or relaxed, and the lattice mismatch at the interface is accommodated by the formation of misfit dislocations. These misfit dislocations generally lie in the plane of the interface, as shown in Fig. 22.1, but dislocations can also thread vertically through the $\text{Si}_{1-x}\text{Ge}_x$ layer.

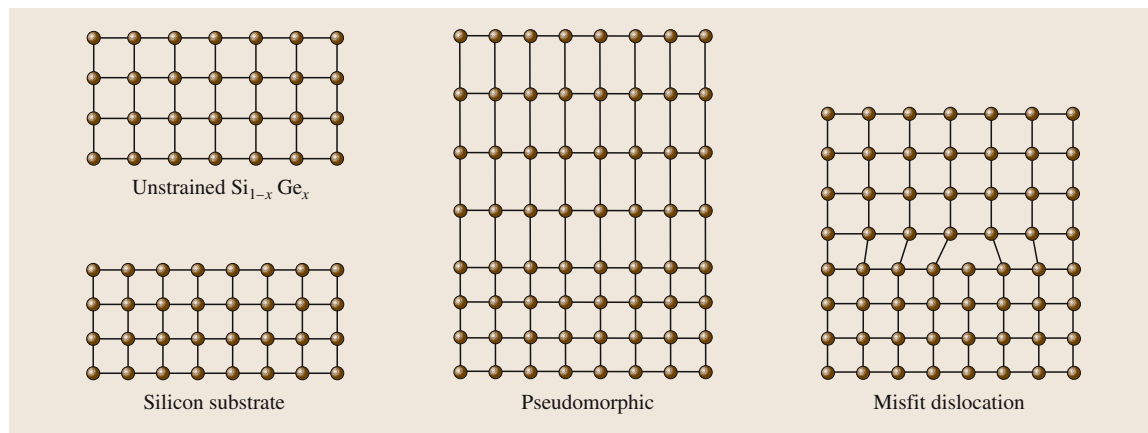


Fig. 22.1 Schematic illustration of pseudomorphic $\text{Si}_{1-x}\text{Ge}_x$ growth and misfit-dislocation formation

22.1.1 Critical Thickness

There is a maximum thickness of $\text{Si}_{1-x}\text{Ge}_x$ that can be grown before relaxation of the strain occurs through the formation of misfit dislocations. This is known as the critical thickness of the $\text{Si}_{1-x}\text{Ge}_x$ layer, and depends strongly on the germanium content, as shown in Fig. 22.2. The original calculations of critical layer thickness were made by *Matthews and Blakeslee* [22.11, 12] on the basis of the mechanical equilibrium of an existing threading dislocation. However, measurements of dislocation densities in $\text{Si}_{1-x}\text{Ge}_x$ showed, in many cases, no evidence of misfit dislocations for $\text{Si}_{1-x}\text{Ge}_x$ layers considerably thicker than the Matthews–Blakeslee limit. These results were explained by *People and Bean* [22.13] who calculated the critical thickness on the assumption that misfit-dislocation generation was determined solely by energy balance. The discrepancy between these two types of calculation can be explained by the observation that strain relaxation in $\text{Si}_{1-x}\text{Ge}_x$ layers occurs gradually. Layers above the People–Bean curve can be considered to be completely relaxed, whereas layers below the Matthews–Blakeslee curve can be considered to be fully strained. These fully strained layers are termed *stable* and will not relax during any subsequent high temperature processing. Layers lying between the two curves are termed *metastable*; these layers may be free of dislocations after growth, but are susceptible to relaxation during later high-temperature processing.

In practice, a number of additional factors influence the critical thickness of a $\text{Si}_{1-x}\text{Ge}_x$ layer. Of

particular importance to both $\text{Si}_{1-x}\text{Ge}_x$ HBTs and $\text{Si}_{1-x}\text{Ge}_x$ MOSFETs, is the effect of a silicon cap layer, which has been shown to increase the critical thickness of the underlying $\text{Si}_{1-x}\text{Ge}_x$ layer. Figure 22.3 shows a comparison of the calculated critical thickness as a function of germanium percentage for stable $\text{Si}_{1-x}\text{Ge}_x$ layers with and without a silicon cap. It can be seen that the critical thickness is more than doubled by the presence of the silicon cap.

The presence of misfit dislocations in devices is highly undesirable, since they create generation/recombination centres, which degrade leakage currents when they are present in the depletion regions of devices. Threading dislocations are also highly undesirable, as they can lead to the formation of emitter/collector pipes in $\text{Si}_{1-x}\text{Ge}_x$ HBTs. When designing $\text{Si}_{1-x}\text{Ge}_x$ devices, it is important that the $\text{Si}_{1-x}\text{Ge}_x$ thickness is chosen to give a stable layer, so that dislocation formation is avoided. A base thickness below the silicon cap curve in Fig. 22.3 will ensure a stable layer, which will withstand ion implantation and high-temperature annealing without encountering problems of relaxation and misfit-dislocation generation.

Considerable research has been done on the oxidation of $\text{Si}_{1-x}\text{Ge}_x$ [22.14, 15], and it has been found that the germanium in the $\text{Si}_{1-x}\text{Ge}_x$ layer does not oxidise, but piles up at the oxide/ $\text{Si}_{1-x}\text{Ge}_x$ interface. This pile-up of germanium makes it difficult to achieve low values of interface state density in oxidised $\text{Si}_{1-x}\text{Ge}_x$ layers. It is therefore advisable to avoid direct oxidation of $\text{Si}_{1-x}\text{Ge}_x$ layers, particularly in $\text{Si}_{1-x}\text{Ge}_x$ MOS technologies. In $\text{Si}_{1-x}\text{Ge}_x$ MOSFETs, a silicon cap is often

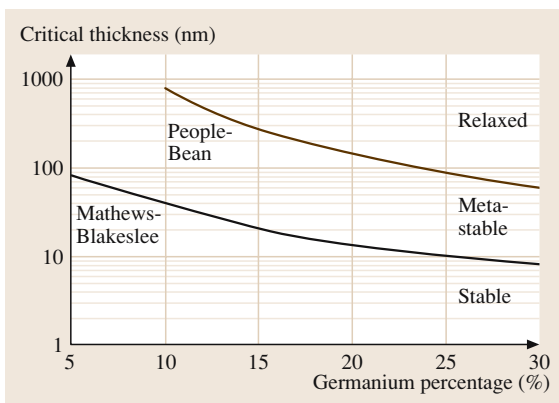


Fig. 22.2 Critical $\text{Si}_{1-x}\text{Ge}_x$ thickness as a function of germanium percentage (after *Iyer et al.* [22.2], copyright 1989 IEEE)

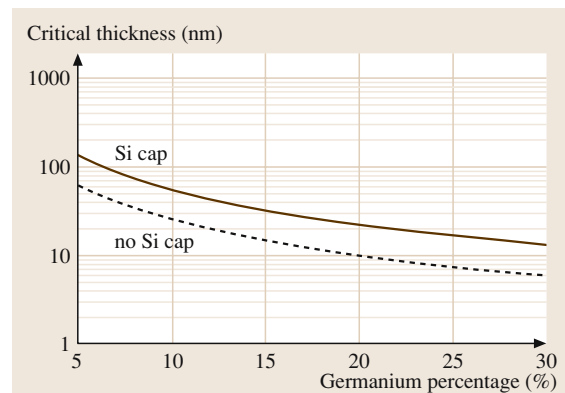


Fig. 22.3 Critical thickness as a function of germanium percentage for stable $\text{Si}_{1-x}\text{Ge}_x$ layers with and without a silicon cap (after *Jain et al.* [22.10], copyright 1992 Elsevier)

included above the $\text{Si}_{1-x}\text{Ge}_x$ layer that can be oxidised to create the gate oxide.

22.1.2 Band Structure

$\text{Si}_{1-x}\text{Ge}_x$ alloys have a smaller band gap than silicon partly because of the larger lattice constant and partly because of the strain. Figure 22.4 shows the variation of band gap with germanium percentage for strained and unstrained $\text{Si}_{1-x}\text{Ge}_x$. It can be seen that the strain has a dramatic effect on the band gap of $\text{Si}_{1-x}\text{Ge}_x$. For 10% germanium, the reduction in band gap compared with silicon is 92 meV for strained $\text{Si}_{1-x}\text{Ge}_x$, compared with 50 meV for unstrained $\text{Si}_{1-x}\text{Ge}_x$. The variation of band gap with germanium content for strained $\text{Si}_{1-x}\text{Ge}_x$ can be described by the following empirical equation:

$$E_G(x) = 1.17 + 0.96x - 0.43x^2 + 0.17x^3 \quad (22.2)$$

The band alignment for compressively strained $\text{Si}_{1-x}\text{Ge}_x$ on unstrained silicon is illustrated schematically in Fig. 22.5. This band alignment is referred to as *type I*, and the majority of the band offset at the hetero-

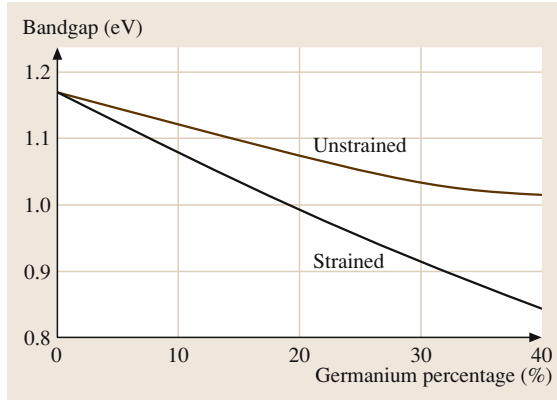


Fig. 22.4 Band gap as a function of germanium percentage for strained [22.16] and unstrained [22.1] $\text{Si}_{1-x}\text{Ge}_x$ (after Iyer et al. [22.2], copyright 1989 IEEE)

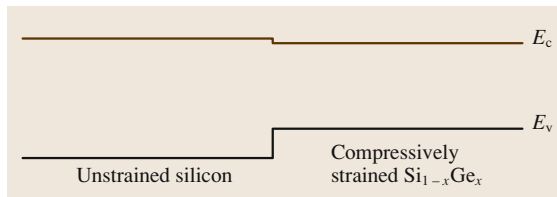


Fig. 22.5 Schematic illustration of the band alignment obtained for a compressively strained $\text{Si}_{1-x}\text{Ge}_x$ layer grown on an unstrained silicon substrate

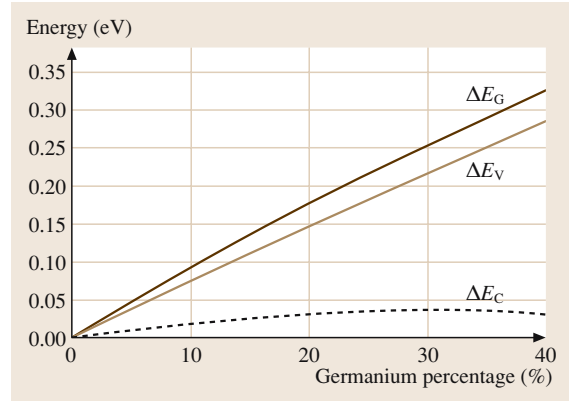


Fig. 22.6 Valence- and conduction-band offsets as a function of germanium percentage for strained $\text{Si}_{1-x}\text{Ge}_x$ grown on unstrained silicon (after Poortmans et al. [22.17], copyright 1993 Elsevier)

junction interface occurs in the valence band, with only a small offset in the conduction band. Different band alignments can be obtained by engineering the strain in the substrate and the grown layer in different ways. For example, *type II* band alignments can be obtained by growing tensile-strained silicon on top of unstrained $\text{Si}_{1-x}\text{Ge}_x$. This arrangement gives large conduction- and valence-band offsets and is used in strained silicon MOSFETs.

Figure 22.6 shows the variation of valence-band offset, ΔE_V , conduction-band offset, ΔE_C , and band-gap narrowing, ΔE_G , with germanium content. It can be seen that the majority of the band offset occurs in the valence band. For example for 10% germanium, the valence-band offset is 0.073 eV, compared with 0.019 eV for the conduction-band offset. The conduction-band offset can therefore be neglected for most practical purposes.

22.1.3 Dielectric Constant

The dielectric constant of $\text{Si}_{1-x}\text{Ge}_x$ can be obtained by linear interpolation between the known values for silicon and germanium [22.17] using the following equation:

$$\epsilon(x) = 11.9(1 + 0.35x) \quad (22.3)$$

22.1.4 Density of States

While, the density of states in the conduction band in $\text{Si}_{1-x}\text{Ge}_x$ is generally assumed to be the same as that in silicon, there is some evidence in the literature to suggest that the density of states in the valence band

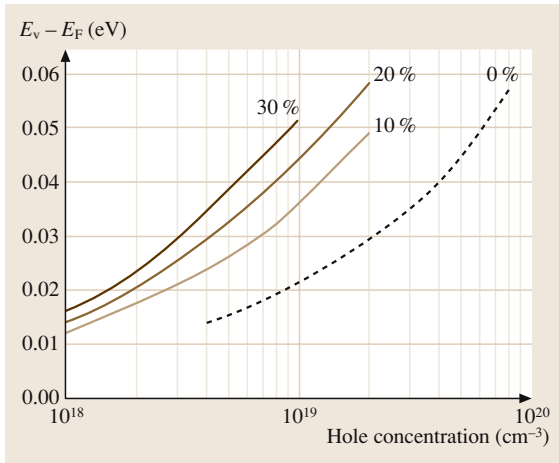


Fig. 22.7 Fermi-level position as a function of hole concentration for $\text{Si}_{1-x}\text{Ge}_x$ with four different germanium concentrations (after Iyer et al. [22.2], copyright 1989 IEEE)

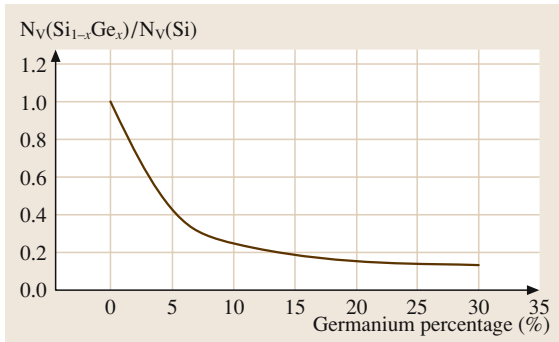


Fig. 22.8 Ratio of density of states in the valence band for $\text{Si}_{1-x}\text{Ge}_x$ to that for Si as a function of germanium percentage (after Poortmans [22.20], copyright 1993 University of Leuven)

is considerably smaller. Manku and Nathan [22.18, 19] have calculated the E - k diagram for strained $\text{Si}_{1-x}\text{Ge}_x$ and shown that the density-of-states hole mass is significantly lower, by a factor of approximately three at 30% germanium. There is some experimental evidence to support this calculation. For example, freeze-out of holes in p-type $\text{Si}_{1-x}\text{Ge}_x$ has been reported to occur at higher temperatures than in p-type silicon [22.20] and enhancements in the majority-carrier, hole mobility have been reported for p-type $\text{Si}_{1-x}\text{Ge}_x$ [22.21].

Using the calculated values of hole density of states of Manku and Nathan [22.18, 19], the hole concentration can be calculated as a function of Fermi-level position.

These results are shown in Fig. 22.7 for $\text{Si}_{1-x}\text{Ge}_x$ with four different germanium contents. It can be seen that the Fermi level moves deeper into the valence band as the germanium concentration increases. Figure 22.8 shows the ratio of the calculated density of states in the valence

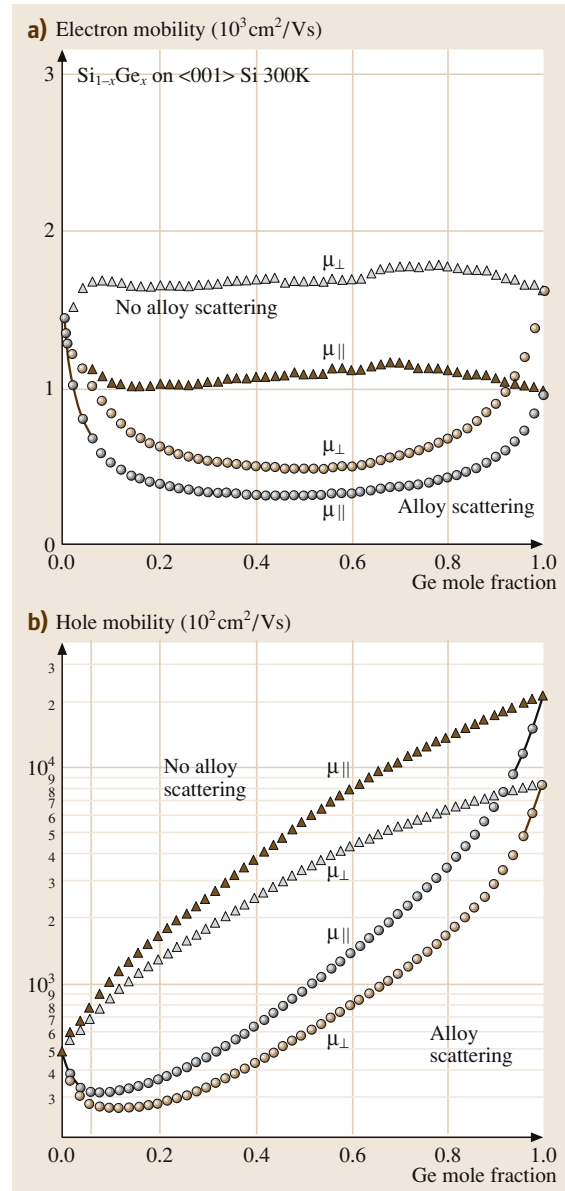


Fig. 22.9a,b Calculated 300-K electron (a) and hole (b) in-plane and out-of-plane low-field mobilities in strained $\text{Si}_{1-x}\text{Ge}_x$ grown on (100) Si (after Fischetti et al. [22.22], copyright 1996 American Institute of Physics)

band for $\text{Si}_{1-x}\text{Ge}_x$ to that for silicon as a function of germanium content. It is clear that the density of states in the valence band for $\text{Si}_{1-x}\text{Ge}_x$ is significantly lower than that for silicon at germanium contents of practical interest.

22.1.5 Majority-Carrier Mobility in Strained $\text{Si}_{1-x}\text{Ge}_x$

Values of in-plane and out-of-plane low-field mobility in strained $\text{Si}_{1-x}\text{Ge}_x$ grown on (100) Si have been calculated by Fischetti et al. [22.22], and are shown in Fig. 22.9. There is some uncertainty in the chosen values of alloy scattering parameters used in the calculations, but nevertheless the results are representative of current understanding. These results show a large enhancement of low-field hole mobility for $\text{Si}_{1-x}\text{Ge}_x$ compared with unstrained silicon, but only a modest enhancement of low-field electron mobility. These results indicate that strained $\text{Si}_{1-x}\text{Ge}_x$ channels can be used to significantly improve the mobility of p-channel MOSFETs, but little benefit is obtained for n-channel MOSFETs. For this reason, industry focus has moved away from channels realised in $\text{Si}_{1-x}\text{Ge}_x$ to channels realised in tensile-strained silicon, as discussed below.

22.1.6 Majority-Carrier Mobility in Tensile-Strained Si on Relaxed $\text{Si}_{1-x}\text{Ge}_x$

Tensile-strained silicon can be produced by growing a thin silicon layer on top of a relaxed $\text{Si}_{1-x}\text{Ge}_x$ virtual substrate. Figure 22.10 shows a typical virtual substrate for a surface-channel MOS transistor. A graded $\text{Si}_{1-x}\text{Ge}_x$ layer is grown on top of the silicon substrate with the Ge content varying from 0 to 30%. Misfit dislocations will form in this layer, but the majority of dislocations will be in the plane of the $\text{Si}_{1-x}\text{Ge}_x$ layer and only a small percentage will propagate vertically

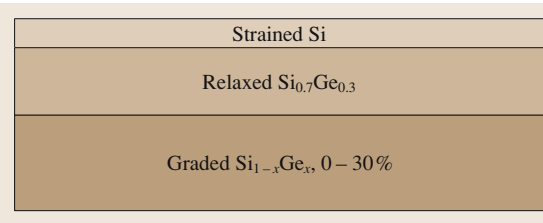


Fig. 22.10 Schematic illustration of a typical tensile-strained Si layer grown on top of a $\text{Si}_{1-x}\text{Ge}_x$ virtual substrate

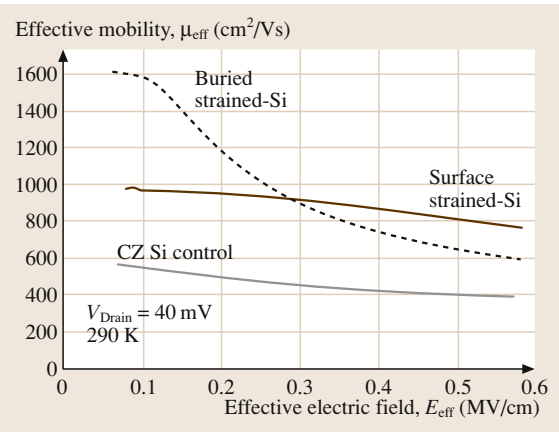


Fig. 22.11 Effective electron mobility as a function of effective electric field for strained Si MOSFETs fabricated on a 30% $\text{Si}_{1-x}\text{Ge}_x$ virtual substrate (after Welser et al. [22.23], copyright 1994 IEEE)

to the surface of the layer. A 30%-relaxed $\text{Si}_{1-x}\text{Ge}_x$ buffer is then grown followed by a thin tensile-strained $\text{Si}_{1-x}\text{Ge}_x$ layer in which the channel is fabricated. The key to any virtual substrate growth process is the minimisation of dislocation propagation to the surface of the wafer.

Figure 22.11 shows typical values of effective electron mobility obtained from measurements on n-channel MOS transistors for a $\text{Si}_{1-x}\text{Ge}_x$ virtual substrate with 30% Ge [22.23]. For the surface-channel strained Si device, the effective mobility is enhanced by 80% compared with the Si control transistor due to the tensile strain in the $\text{Si}_{1-x}\text{Ge}_x$ layer.

Enhanced hole mobility can also be obtained in tensile-strained Si grown on a $\text{Si}_{1-x}\text{Ge}_x$ virtual substrate, though higher germanium contents are needed to obtain a significant mobility enhancement. Figure 22.12 shows typical values of effective hole mobility in strained Si for Ge contents between 35 and 50% [22.24]. The effective mobility of the strained Si device is enhanced by 100% compared with the Si control device.

22.1.7 Minority-Carrier Mobility in Strained $\text{Si}_{1-x}\text{Ge}_x$

$\text{Si}_{1-x}\text{Ge}_x$ HBTs are minority-carrier devices and hence values of the minority-carrier mobility of more interest than the majority-carrier mobility. Unfortunately very few measurements of minority-carrier mobility have been made in $\text{Si}_{1-x}\text{Ge}_x$. Poortmans [22.20] inferred values of minority-carrier mobility from measure-

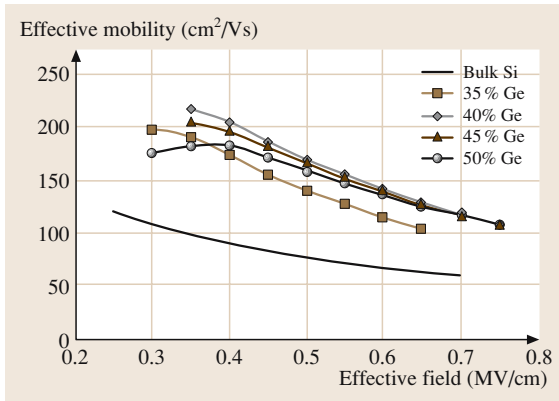


Fig. 22.12 Effective hole mobility as a function of effective electric field for strained Si MOSFETs fabricated on a $\text{Si}_{1-x}\text{Ge}_x$ virtual substrate with Ge contents in the range 35–50% (after Leitz et al. [22.24], copyright 2002 IEEE)

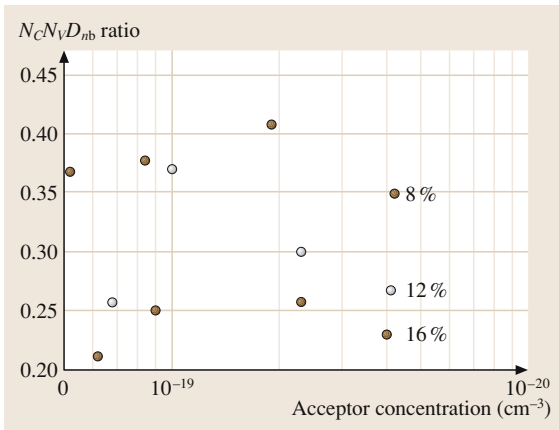


Fig. 22.13 Measured values of the ratio of $N_C N_V D_{nb}$ in $\text{Si}_{1-x}\text{Ge}_x$ to that in Si as a function of acceptor concentration (after Poortmans [22.20], copyright 1993 University of Leuven)

ments on $\text{Si}_{1-x}\text{Ge}_x$ HBTs and found an enhancement in mobility compared with silicon by a factor of 1.2–1.4 for base doping concentrations in the range 5×10^{18} – $5 \times 10^{19} \text{ cm}^{-3}$. Given the scarcity of measured data on minority-carrier mobility and density of states in $\text{Si}_{1-x}\text{Ge}_x$, the most reliable way of calculating the expected gain improvement in a $\text{Si}_{1-x}\text{Ge}_x$ HBT is to use

data directly obtained from measurements on $\text{Si}_{1-x}\text{Ge}_x$ HBTs. The gain enhancement in a $\text{Si}_{1-x}\text{Ge}_x$ HBT is determined by the ratio of the product $N_C N_V D_{nb}$ in $\text{Si}_{1-x}\text{Ge}_x$ and Si, together with the band-gap narrowing due to the strained $\text{Si}_{1-x}\text{Ge}_x$. Figure 22.13 shows a graph of this $N_C N_V D_{nb}$ ratio as a function of acceptor concentration for three values of germanium content. It can be seen that for germanium contents of practical interest, in the range 11–16%, this ratio has a value of around 0.25.

22.1.8 Apparent Band-Gap Narrowing in $\text{Si}_{1-x}\text{Ge}_x$ HBTs

In $\text{Si}_{1-x}\text{Ge}_x$ HBTs, the apparent band-gap narrowing is often quoted, which combines the effect of the band-gap reduction and the effect of high doping. Poortmans et al. [22.17] have developed a theoretical approach that has been shown to be in reasonable agreement with experiment. Figure 22.14 shows the apparent band-gap narrowing in $\text{Si}_{1-x}\text{Ge}_x$ as a function of acceptor concentration for three values of germanium content. At low acceptor concentrations, the apparent band-gap narrowing in $\text{Si}_{1-x}\text{Ge}_x$ is slightly higher than that in silicon, but at acceptor concentrations in the range 1 – $2 \times 10^{19} \text{ cm}^{-3}$, the apparent band-gap narrowing is approximately the same. This latter doping range is the base doping range that is of practical interest for $\text{Si}_{1-x}\text{Ge}_x$ HBTs.

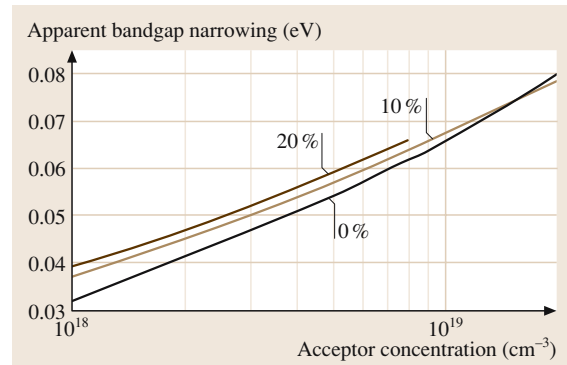


Fig. 22.14 Apparent band-gap narrowing as a function of acceptor concentration for $\text{Si}_{1-x}\text{Ge}_x$ with three different germanium concentrations (after Poortmans et al. [22.17], copyright 1993 Elsevier)

22.2 Optical Properties of SiGe

Interest in the optical properties of SiGe stems from the desire to design silicon-based optoelectronic devices as well as the usefulness of many optical techniques in material analysis. The optical properties of bulk SiGe provides an important starting point for any attempt at in-depth understanding, however, nearly all real applications of SiGe involve the use of thin, strained layers. Beyond this, the formation of SiGe quantum structures within silicon devices has remained one of the most promising methods by which device engineers hope to improve the largely unimpressive optical behaviour of silicon that is brought about by its indirect band gap. The growth of quantum wells, quantum wires and quantum dots in the Si/SiGe system has been extensively explored [22.26–31]. However, there has as yet been little success at using Si/SiGe quantum structures to produce efficient silicon-based light emitters, although there have been impressive attempts [22.31]. Perhaps the most promising devices based on SiGe quantum wells are near-infra-red photodetectors in which the SiGe can be used to enhance sensitivity at the optical-communications wavelengths [22.32–36]. More

futuristic applications of SiGe quantum wells include devices based on transitions between the confined energy levels in quantum wells. Devices based on these *inter-subband* transitions include quantum-well infrared photodetectors (QWIPs) [22.37, 38] and quantum cascade lasers [22.39, 40].

22.2.1 Dielectric Functions and Interband Transitions

A range of 1- μm -thick $\text{Si}_{1-x}\text{Ge}_x$ films grown by MBE on Si(100) substrates have been studied by spectroscopic ellipsometry to yield their complex dielectric functions at room temperature [22.25]. Both the real (ϵ_1) and imaginary (ϵ_2) parts of the measured dielectric function for Ge compositions of 0, 0.2, 0.4, 0.6, 0.8 and 1.0 are shown in Fig. 22.15a and b, respectively.

In Fig. 22.15b the absorption structures observed at 3.4 and 4.2 eV in the spectra shown for silicon originate from direct band-to-band transitions at various regions in the Brillouin zone of silicon. The structure seen around 3.4 eV is due to E'_0 , E_1 and $E_1 + \Delta_1$ interband transition

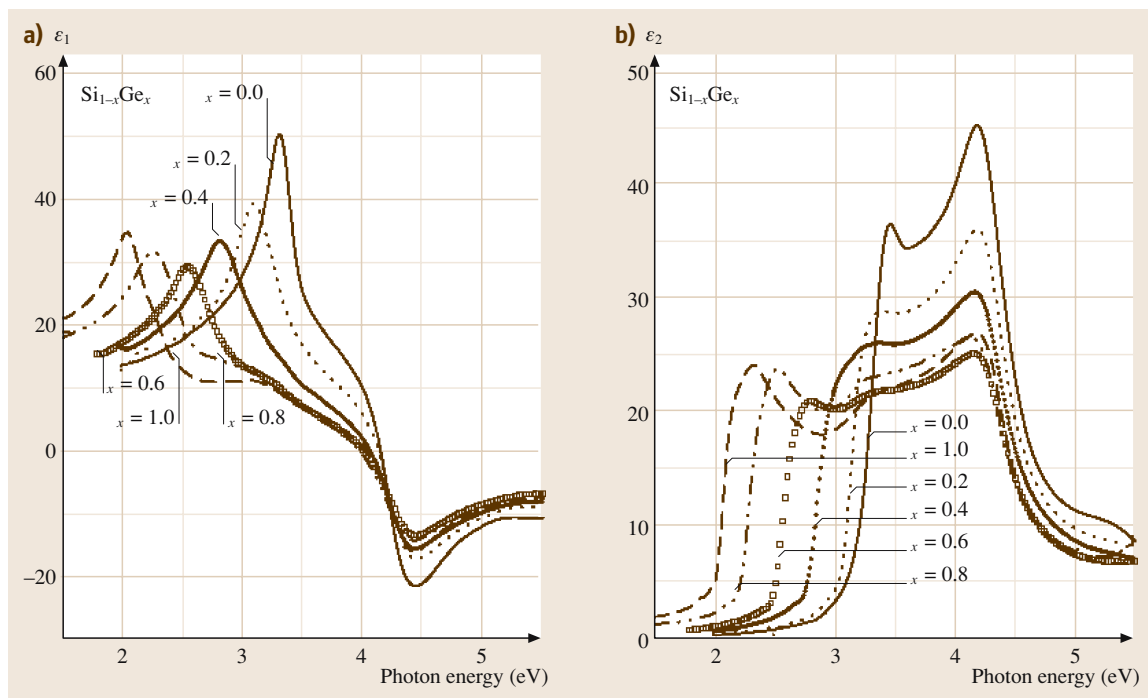


Fig. 22.15a,b Real (a) and imaginary (b) parts of the dielectric functions of relaxed $\text{Si}_{1-x}\text{Ge}_x$ alloys with composition x indicated in the legend (after Bahng et al. [22.25], copyright 2001, American Physical Society)

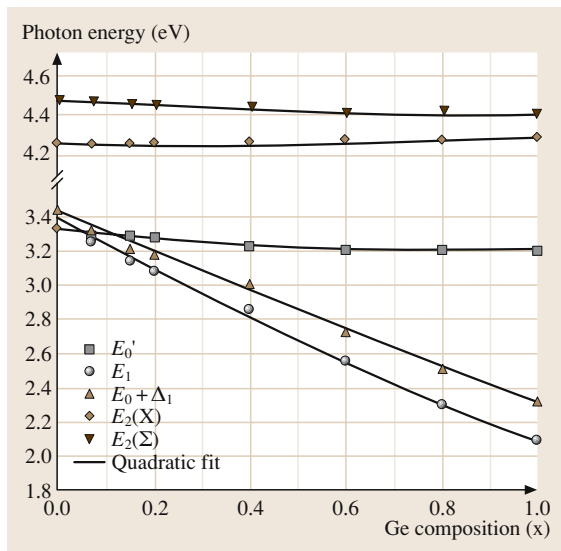


Fig. 22.16 Evolution of E'_0 , E_1 , $E_1 + \Delta_1$, $E_2(X)$ and $E_2(\Sigma)$ transition energies for relaxed $\text{Si}_{1-x}\text{Ge}_x$ with composition x (after Bahng et al. [22.25], copyright 2001, American Physical Society)

edges, whereas the structure at 4.2 eV is due to $E_2(X)$ and $E_2(\Sigma)$ edges [22.25]. The evolution of each of these transition edges for the full range of SiGe compositions is shown in Fig. 22.16.

The quadratic fits shown in Fig. 22.16 are as follows [22.25]:

$$E'_0(x) = 3.337 - 0.348x + 0.222x^2 \quad (22.4)$$

$$E_1(x) = 3.398 - 1.586x + 0.27x^2 \quad (22.5)$$

$$E_1 + \Delta_1(x) = 3.432 - 1.185x + 0.065x^2 \quad (22.6)$$

$$E_2(X)(x) = 4.259 - 0.052x + 0.084x^2 \quad (22.7)$$

$$E_2(\Sigma)(x) = 4.473 - 0.139x + 0.072x^2 \quad (22.8)$$

22.2.2 Photoluminescence

The emission properties of semiconductor structures are of fundamental interest to scientists as well as being an important analytical technique for engineers. In general, the features of low-temperature photoluminescence spectra are very dependent on the specific conditions under which materials are grown and treated. This is because emission energies and emission rates are often sensitive to even small variations of impurity or defect densities, as well as variations in strain or composition. A brief examination of low-temperature

photoluminescence spectra is nearly always sufficient to allow a simple qualitative assessment of material quality; alternatively, detailed analysis can permit a broad range of material or structural parameters to be assessed or determined. No two photoluminescence spectra are the same. In this section and the section that follows on photoluminescence studies of Si/SiGe quantum wells, we will present a range of spectra that represent most of the key features that have been observed.

Weber and Alonso [22.41] have provided a very useful study of the near-band-gap photoluminescence of bulk SiGe alloys. In their study bulk SiGe samples are cut from nominally undoped polycrystalline ingots prepared by a zone-leveiling technique. Figure 22.17 shows the photoluminescence spectra for a range of compositions. Samples were excited using the 514-nm line of an argon ion laser and the sample temperature was 4.2 K.

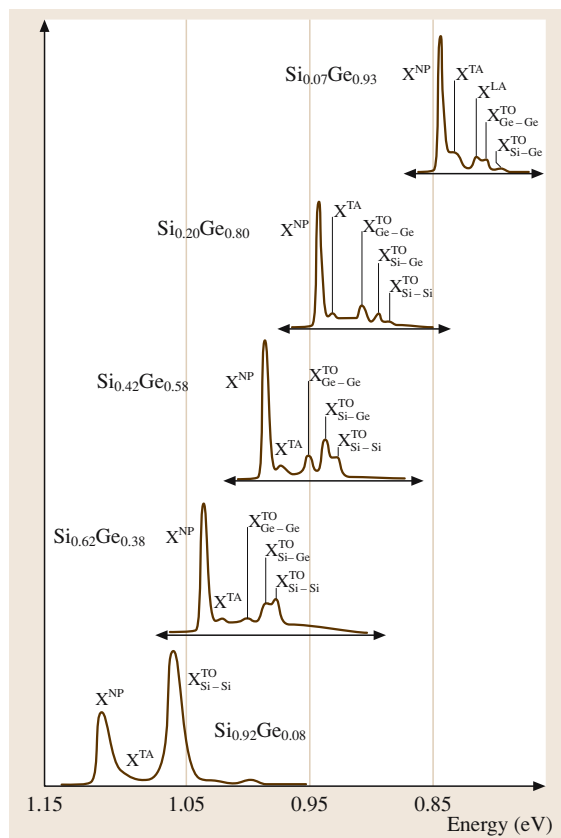


Fig. 22.17 Near-band-gap photoluminescence spectra for several bulk SiGe samples (after Weber et al. [22.41], copyright 1989, American Physical Society)

Excitonic emission lines are a strong feature of photoluminescence spectra when the thermal energy of the semiconductor is less than the exciton binding energy. Each spectrum featured in Fig. 22.17, across the full range of SiGe compositions, show similar *excitonic* features.

In most spectra the most pronounced peak is the no-phonon (X^{NP}) line caused by the optical recombination of excitons bound to shallow impurities. In the case of the no-phonon line, momentum is conserved through interaction with the binding impurity. There are many candidate atoms for these shallow impurities and with B, P and As having binding energies of 4.2, 5.0 and 5.6 meV, respectively [22.42]. These bound energy states will tend to dominate luminescence spectra for doped samples and will always tend to be present in nominally undoped samples. The no-phonon line is accompanied by transverse-optical (X^{TO}) or transverse acoustic (X^{TA}) phonon replicas that are created as photon emission is accompanied by the momentum-conserving creation of lattice vibrations in Si–Si, Si–Ge or Ge–Ge bonds.

Figure 22.18 shows how the spectrum from a bulk $Si_{0.915}Ge_{0.085}$ sample develops with increasing temperature [22.41]. With increasing temperature the X^{NP} line thermalises to leave the free-exciton (FE^{NP}) line. Here, emission from nominally free excitons is greatly enhanced for the alloy samples by local fluctuations in composition that provide momentum-conserving scattering centres [22.43].

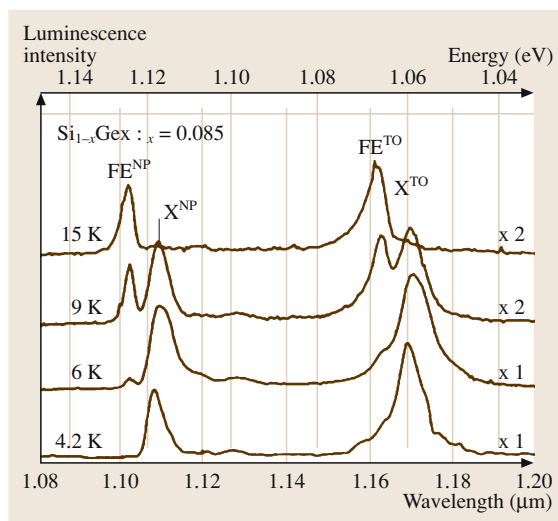


Fig. 22.18 Photoluminescence spectra of a bulk $Si_{0.915}Ge_{0.085}$ sample at different temperatures (after Weber et al. [22.41], copyright 1989, American Physical Society)

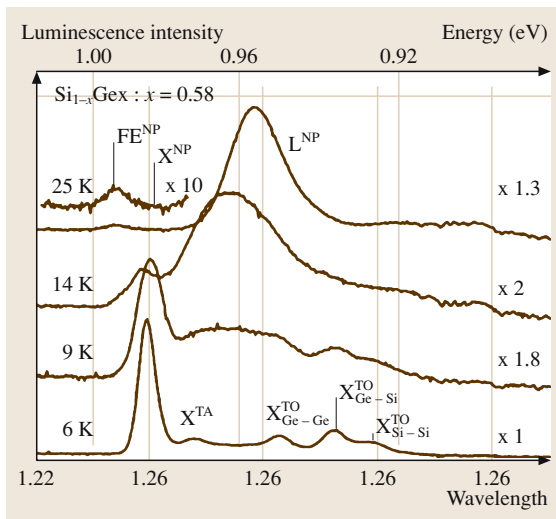


Fig. 22.19 Photoluminescence spectra of a bulk $Si_{0.42}Ge_{0.58}$ sample at different temperatures (after Weber et al. [22.41], copyright 1989, American Physical Society)

At higher temperatures the free-exciton line is also thermalised and all fine structure is lost, at temperatures around 25 K broad luminescence bands are commonly observed (Fig. 22.19 [22.41]). These deep luminescence bands are difficult to assign and have been ascribed to impurities, structural defects and, as in the case of the line presented in Fig. 22.19, potential wells formed by alloy fluctuations [22.41].

Weber and Alonso [22.41] use their data to provide analytical expressions for both the X^{NP} and L bands for bulk $Si_{1-x}Ge_x$ in the range $0 \leq x \leq 0.85$ as follows:

$$E_{gx}^{(x)}(x) = 1.155 - 0.43x + 0.206x^2 \text{ eV} \quad (22.9)$$

$$E_{gx}^L(x) = 2.010 - 1.270x \text{ eV} \quad (22.10)$$

At low temperatures narrow excitonic luminescence features are indicative of defect-free material, and in this way low-temperature photoluminescence becomes a good qualitative tool with which material quality can be assessed.

22.2.3 SiGe Quantum Wells

Figure 22.20 shows the first excitonic luminescence spectra from a Si/SiGe multiple quantum well grown by atmospheric-pressure CVD [22.28]. As we can see, many of the features seen in the photoluminescence spectra of quantum-well samples are similar to

Fig. 22.20 Excitonic photoluminescence spectrum of a SiGe quantum well (after Grutzmacher et al. [22.28], copyright 1993, AVS)

those seen from the bulk samples described in the previous section. Again, the most pronounced peak is the no-phonon (NP) line and this is accompanied by phonon replicas, including, impressively, a two-phonon replica of the NP line ($\text{TO} + \text{TO}^{\text{Si-Si}}$). The most significant difference between bulk and quantum well spectra is the energy positions of the excitonic features as these are shifted by quantum confinement effects.

Robbins et al. [22.26], have provided one of the most detailed studies of near-band-gap photoluminescence from pseudomorphic SiGe layers and provide analytical expressions for all factors pertaining to the energy positions of the excitonic energy gap for $\text{Si}_{1-x}\text{Ge}_x$ quantum wells in the range $0 < x < 0.24$. The effects of alloying, confinement, band offsets, alignment type and exciton binding energy are all taken into account. Samples used in the study were grown by low-pressure CVD at 920 °C; a typical set of photoluminescence spectra are shown in Fig. 22.21.

The exciton band gap at 4.2 K is considered for thick (50-nm) strained layers (E_X^S) where the energies are not affected by quantum shifts, the following expression is derived:

$$E_X^S(x) = 1.155 - 0.874x + 0.376x^2 \text{ eV}, \quad (x < 0.25) \quad (22.11)$$

Here the presence of strain is responsible for the differences from the expression obtained for bulk samples (22.9). An expression for the exciton binding energy $E_B^C(x)$ is theoretically derived for the cubic alloy and the following quadratic expression is fitted:

$$E_B^C(x) \approx 0.0145 - 0.022x + 0.020x^2 \text{ eV}, \quad (x < 0.25) \quad (22.12)$$

A strain-corrected expression is also provided but this is found to modify (22.12) only slightly. Thus by adding (22.11) and (22.12) an expression for the band gap can be obtained.

$$E_C - E_V \approx 1.17 - 0.896x + 0.39x^2 \text{ eV}, \quad (x < 0.25) \quad (22.13)$$

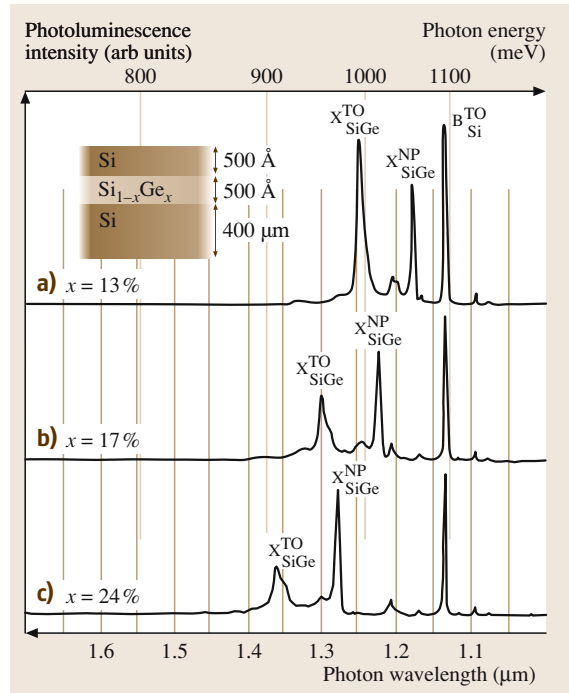
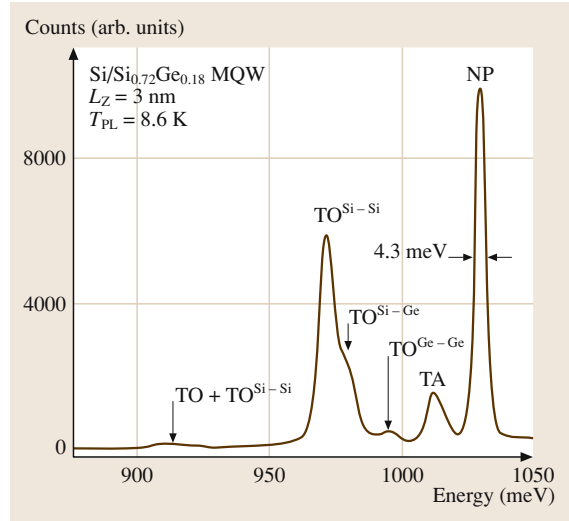


Fig. 22.21 4.2-K photoluminescence spectra from layers with the nominal structure shown in the inset (514-nm Ar-ion laser excitation) (after Robbins et al. [22.26], copyright 1992, American Physical Society)

22.3 Growth of Silicon–Germanium

Over the past ten years and more there have been rapid developments in techniques for the growth of Si and $\text{Si}_{1-x}\text{Ge}_x$ epitaxial layers at low temperatures. This has been made possible by a number of changes in the design of epitaxy equipment and by improvements to growth processes. There are two main prerequisites for the growth of epitaxial layers at low temperature:

- Establishment of a clean surface prior to growth [22.44–47]
- Growth in an ultra-clean environment [22.48–50]

The removal of oxygen and carbon is the main problem in establishing a clean surface prior to growth. A clean silicon surface is highly reactive and oxidises in air even at room temperature. The secret of low-temperature epitaxial growth is therefore the removal of this native oxide layer and the maintenance of a clean surface until epitaxy can begin. Two alternative approaches to pre-epitaxy surface cleaning have been developed, as described below.

22.3.1 In-Situ Hydrogen Bake

The concept that underlies this surface clean is the controlled growth of a thin surface oxide layer, followed by its removal in the epitaxy reactor using a hydrogen bake. The controlled growth of the surface oxide layer is generally achieved using a Radio Corporation of America (RCA) clean [22.44] or a variant [22.45]. The oxide created by the RCA clean is removed in the reactor using an in-situ bake in hydrogen for around 15 min at a temperature in the range 900–950 °C. The temperature required to remove the native oxide depends on the thickness of the oxide, which is determined by the severity of the surface clean.

22.3.2 Hydrogen Passivation

An alternative approach to pre-epitaxy cleaning is to create an oxide-free surface using an ex-situ clean and then move quickly to epitaxial growth before the native oxide can grow. The aim of the ex-situ clean is to produce a surface that is passivated by hydrogen atoms bonded to dangling bonds from silicon atoms on the surface. When the wafers are transferred in the epitaxy reactor, the hydrogen can be released from the surface of the silicon very quickly using a low-temperature bake or even in the early stages of epitaxy without any bake. Meyerson [22.46] has reported that hydrogen desorbs

at 600 °C at a rate of a few monolayers per second, so the hydrogen passivation approach allows epitaxial layers to be grown at low temperatures without the need for a high-temperature bake. The hydrogen-passivated surface is stable for typically 30 min after completion of the ex-situ cleaning [22.47].

22.3.3 Ultra-Clean Epitaxy Systems

Having produced a clean hydrogen-passivated silicon surface, it is clearly important to maintain the state of this surface in the epitaxy system. This necessitates the use of low-pressure epitaxy systems if epitaxial growth at low temperatures is required. Figure 22.22 summarises the partial pressures of oxygen and water vapour that need to be achieved in an epitaxy system if an oxide-free surface is to be maintained at a given temperature [22.48, 49]. This figure shows that epitaxial growth at low temperature requires low partial pressures of oxygen and water vapour, which of course can be achieved by reducing the pressure in the epitaxy system. Research [22.50] has shown that a pressure below 30 Torr is needed to achieve silicon epitaxial growth below 900 °C.

22.3.4 $\text{Si}_{1-x}\text{Ge}_x$ Epitaxy

The growth of $\text{Si}_{1-x}\text{Ge}_x$ epitaxial layers can be achieved over a wide range of temperatures using low-pressure chemical vapour deposition (LPCVD) [22.50] or ultra-high-vacuum chemical vapour deposition (UHV-CVD) [22.51, 52]. The gas used to introduce the germanium into the layers is germane, GeH_4 . The influence of germanium on the growth rate is complex, as illustrated in Fig. 22.23. At temperatures in the range 577–650 °C a peak in the growth rate is seen. At low germanium contents, the growth rate increases with germanium content, whereas at high germanium content, the growth rate decreases with germanium content. In the low-temperature regime it has been proposed that hydrogen desorption from the surface is the rate-limiting step. In $\text{Si}_{1-x}\text{Ge}_x$ this occurs more easily at germanium sites than at silicon sites and hence the growth rate increases with germanium content [22.37]. As the germanium content increases, the surface contains more and more germanium and less and less hydrogen. The rate-limiting step then becomes the adsorption of germane or silane. Robbins [22.53] proposed that the sticking coefficient for germane or silane was lower at germanium sites. This would slow the adsorption rate as the ger-

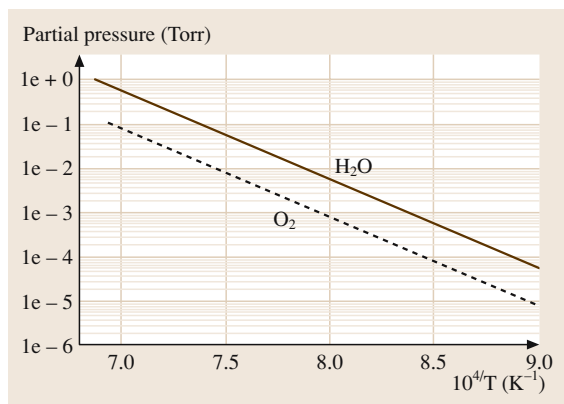


Fig. 22.22 Conditions for oxide formation in an epitaxy system. Note that 1 atm = 1.113 bar = 760 Torr = 1.113×10^5 Pa (after Smith and Ghidini [22.48, 49], copyright Electrochemical Society 1984)

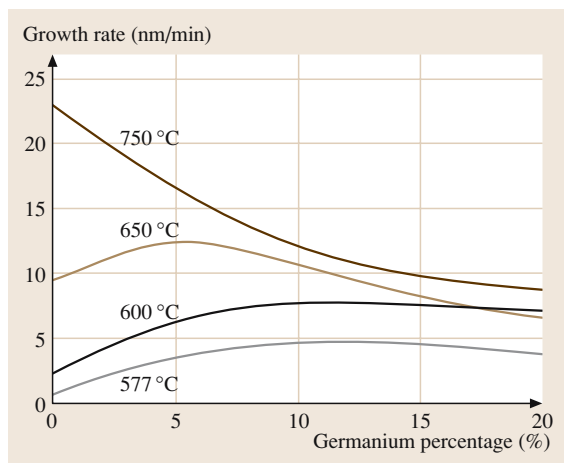


Fig. 22.23 Growth rate of $\text{Si}_{1-x}\text{Ge}_x$ as a function of germanium percentage for temperatures in the range 577–750 °C (after Racanelli et al. [22.54], copyright 1990, American Institute of Physics)

manium content increased and hence slow the growth rate.

22.3.5 Selective $\text{Si}_{1-x}\text{Ge}_x$ Epitaxy

Selective epitaxy is the growth of a single-crystal layer in a window, with complete suppression of growth elsewhere, and can be achieved in a number of different ways. The most common method of achieving both selective Si and $\text{Si}_{1-x}\text{Ge}_x$ epitaxy is by introducing chlorine or HCl into the growth chamber. This can either be

done by adding chlorine or HCl as a separate gas or by using a growth gas that contains chlorine, for example dichlorosilane, SiH_2Cl_2 . With chlorine chemistry, selective growth of silicon and $\text{Si}_{1-x}\text{Ge}_x$ can be achieved to both silicon dioxide and silicon nitride.

Chlorine is reported to have two effects that lead to selective growth. First it increases the surface mobility of silicon and germanium atoms, so that atoms deposited on the oxide or nitride layer are able to diffuse across the surface to the window where the growth is occurring. Second it acts as an etch [22.50] and hence can remove silicon or germanium atoms deposited on the oxide or nitride. The strength of the etching action increases with chlorine content and, if the chlorine content is too high, etching of the substrate will occur instead of epitaxial growth.

A typical growth process for selective silicon epitaxy would use silane and a few percent of HCl [22.50]. The growth rate for this process is shown in Fig. 22.24, and compared with the growth rate for dichlorosilane and silane epitaxy. It can be seen that the activation energy for the silane-plus-HCl process is very similar to that for the dichlorosilane process, indicating that the growth mechanisms are similar. One disadvantage of chlorine-based growth processes over the silane process is a lower growth rate at low temperatures, as can clearly be seen in Fig. 22.24. It is also possible to grow silicon selectively using dichlorosilane and HCl [22.55].

Selective $\text{Si}_{1-x}\text{Ge}_x$ growth is generally easier to achieve than selective silicon growth, as illustrated in

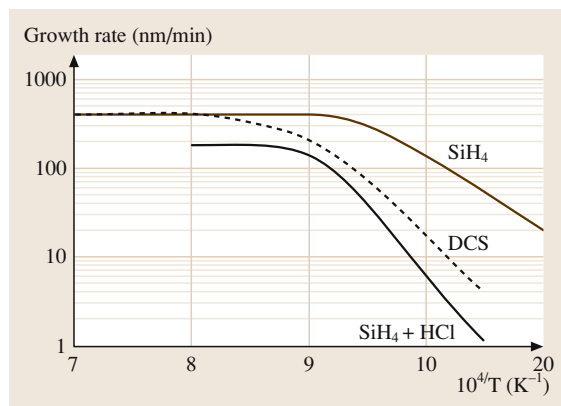


Fig. 22.24 Silicon growth rate as a function of reciprocal temperature for three different growth gases: 40 sccm of SiH_4 , 80 sccm of dichlorosilane and 20 sccm of SiH_4 with 2 sccm of HCl. The hydrogen flow was 2 slm (after Regolini et al. [22.50], copyright 1989 American Institute of Physics)

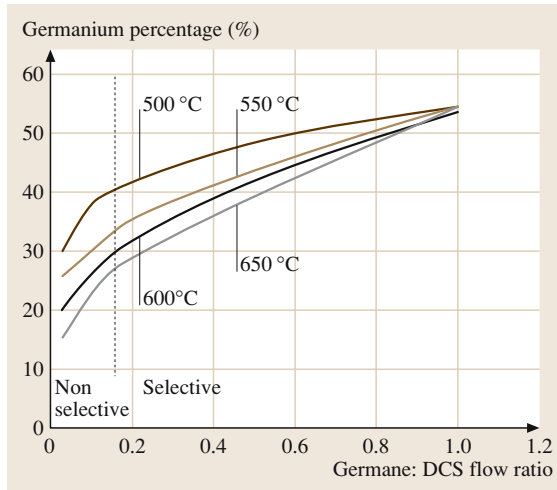


Fig. 22.25 Germanium percentage as a function of germane: dichlorosilane (DCS) flow ratio for temperatures in the range 500–650 °C showing the move from nonselective to selective growth as the proportion of germane in the gas flow increases (after *Zhong et al.* [22.56], copyright 1990, American Institute of Physics)

Fig. 22.25 [22.56] for $\text{Si}_{1-x}\text{Ge}_x$ growth using germane and dichlorosilane. The growth moves from nonselective to selective as the proportion of germane in the gas flow increases.

Arrhenius plots for $\text{Si}_{1-x}\text{Ge}_x$ growth using germane and dichlorosilane are shown in Fig. 22.26 for $\text{Si}_{1-x}\text{Ge}_x$ layers grown using germane and dichlorosilane and for two different HCl flows. It can be seen that the growth rate decreases and the activation energy increases with increasing HCl flow. The explanation proposed for this behaviour is that the limiting growth mechanism changes from hydrogen desorption from the growing surface to chlorine or HCl desorption from the surface [22.57]. This decrease in growth rate at high HCl flows is a disadvantage because it leads

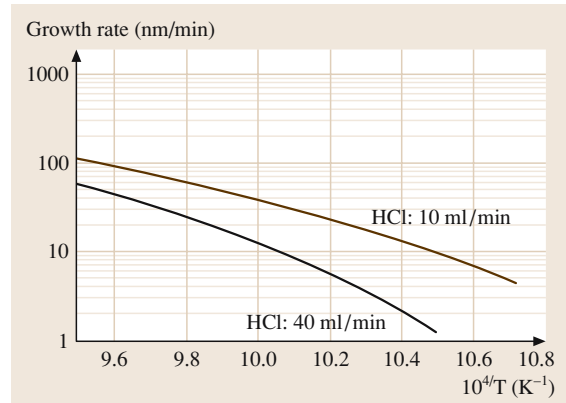


Fig. 22.26 Arrhenius plots for $\text{Si}_{1-x}\text{Ge}_x$ growth at two different HCl flow rates. The dichlorosilane and germane flow rates were fixed at 100 and 8 ml/min respectively (after *Kiyota et al.* [22.57], copyright 2002, IEEE)

to increased growth times. High HCl flows can also cause surface roughening when the $\text{Si}_{1-x}\text{Ge}_x$ layer is heavily boron-doped [22.57]. These considerations demonstrate that the HCl flow should be chosen to be to the smallest value that is consistent with good selective epitaxy.

Silane can be used for selective silicon epitaxy if the growth is performed at a high temperature. This approach relies on the fact that nucleation of growth on oxide is more difficult than that on silicon. This incubation time for growth on an oxide layer is relatively long at high temperatures but much shorter for growth at low temperatures. Selective silicon layers 1 μm thick can be grown using silane at a temperature of 960 °C [22.58], but the achievable layer thickness decreases with decreasing temperature. At 800 °C the maximum selective silicon layer thickness is around 130 nm, at 700 °C it is around 60 nm, and at 620 °C it is around 40 nm. Selective growth to silicon dioxide can be achieved using silane only, but not to silicon nitride.

22.4 Polycrystalline Silicon–Germanium

In the past ten years there has been increasing interest in polycrystalline silicon–germanium for a number of applications that require polycrystalline material deposition at low temperature (around 600 °C). Examples of potential applications are thin film transistors, gates of MOS transistors and polySiGe emitters for SiGe HBTs.

In thin-film transistor technologies [22.59–61], polycrystalline silicon–germanium is compatible with the low-thermal-budget processing that is needed to produce thin-film devices for large-area electronics. The key physical property of polycrystalline silicon–germanium that makes it attractive is its lower melting point than silicon. This means that processes such as deposition,

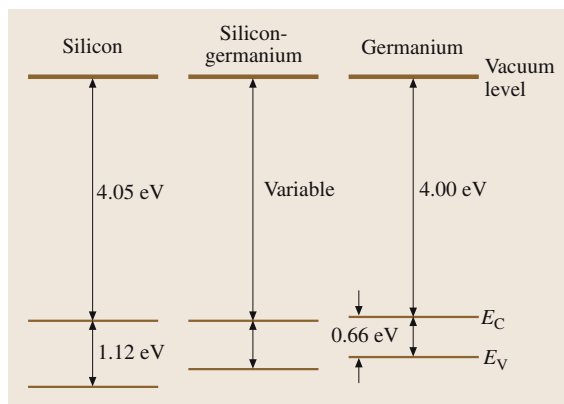


Fig. 22.27 Band-energy levels in silicon, silicon–germanium and germanium

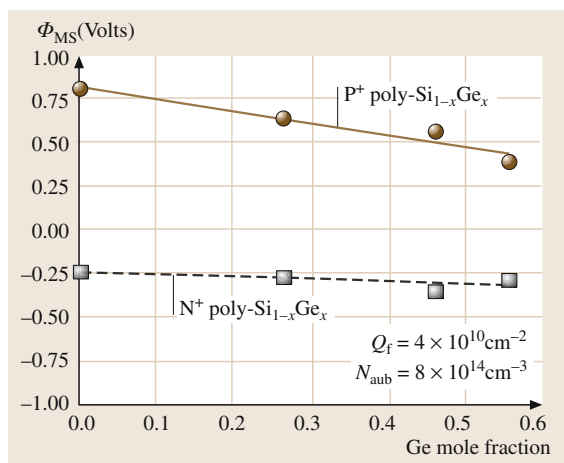


Fig. 22.28 Work-function difference between a polySi_{1-x}Ge_x gate and an n-type substrate as a function of germanium content (after King et al. [22.60], copyright 1994, IEEE)

crystallisation, grain growth and dopant activation will occur at a lower temperature than in silicon. Thus lower temperature processes can be used for polySiGe devices and hence it is preferable to polySi in applications with tight thermal-budget requirements.

In MOS transistors, polycrystalline silicon–germanium is attractive as a gate material for future generations of MOS transistor, since the germanium content in the silicon–germanium layer can be varied by 200–300 mV in the direction of a mid-gap gate [22.63–65]. This can be understood from Fig. 22.27, which compares the conduction- and valence-band energy levels in single-crystal silicon, silicon–germanium and germanium. Silicon and germanium have similar

electron affinities (4.05 and 4.00 eV respectively), but germanium has a much smaller band gap (0.66 eV compared with 1.12 eV). The energy difference between the valence band and the vacuum level is therefore about 0.5 eV smaller in germanium than in silicon. In silicon–germanium, this energy difference can be varied by varying the germanium content. This allows the threshold voltage of p-channel MOS transistors to be tuned by varying the germanium content in the polySi_{1-x}Ge_x gate.

Figure 22.28 shows values of work-function difference between the polySi_{1-x}Ge_x gate and the n-type silicon substrate as a function of germanium content in a polySi_{1-x}Ge_x gate [22.60]. The work function is defined as the difference in energy between the vacuum level and the Fermi level. In p⁺ polySi_{1-x}Ge_x the Fermi level is near the valence band and hence the work-function difference varies strongly with germanium content. In n⁺ polySi_{1-x}Ge_x the Fermi level is near the conduction band and hence the work-function difference varies little with germanium content.

In Si_{1-x}Ge_x HBTs, polySi_{1-x}Ge_x has potential as an emitter of a SiGe HBT [22.62]. In bipolar transistors, the breakdown voltage, BV_{CEO} , is inversely proportional to the gain [22.66] and hence transistors with a high gain have lower values of breakdown voltage. Si_{1-x}Ge_x HBTs inherently have high values of the gain because the reduced band gap of the Si_{1-x}Ge_x base enhances the collector current. The base current is unchanged by the Si_{1-x}Ge_x base and hence the gain, which is the

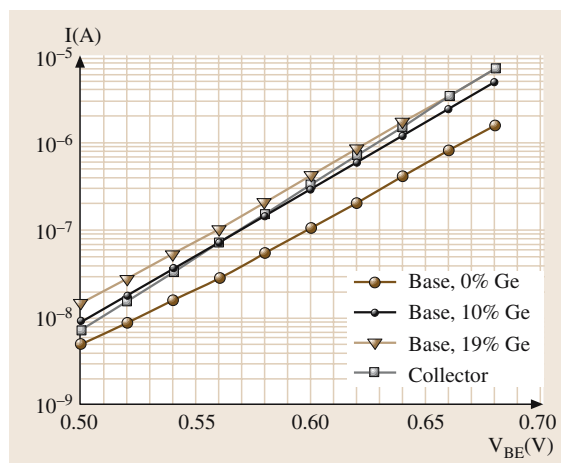


Fig. 22.29 Use of a polySi_{1-x}Ge_x emitter to vary the base current of a Si_{1-x}Ge_x HBT and hence give the best trade-off between gain and breakdown voltage BV_{CEO} . (after Kunz et al. [22.62], copyright 2003, IEEE)

ratio of collector current to base current, is increased. The use of a polySi_{1-x}Ge_x emitter instead of a polySi emitter provides a reduced band gap in the emitter, which enhances the base current, and thereby reduces the gain. Typical measured values of base current in a polySi_{1-x}Ge_x emitter are shown in Fig. 22.29, where it can be seen that 19% germanium gives a factor of approximately four reduction in gain. A polySi_{1-x}Ge_x emitter therefore allows the gain to be tuned to give the best trade-off between gain and breakdown voltage BV_{CEO} .

22.4.1 Electrical Properties of Polycrystalline Si_{1-x}Ge_x

Figure 22.30 shows the sheet resistance as a function of anneal temperature for boron- and phosphorus-doped Si_{1-x}Ge_x for different germanium contents. For boron-doped polySi_{1-x}Ge_x the sheet resistance decreases with increasing germanium content, with the decrease being large between 0 and 25% germanium and smaller between 25 and 50% germanium. In contrast, for phosphorus-doped polySi_{1-x}Ge_x the sheet resistance

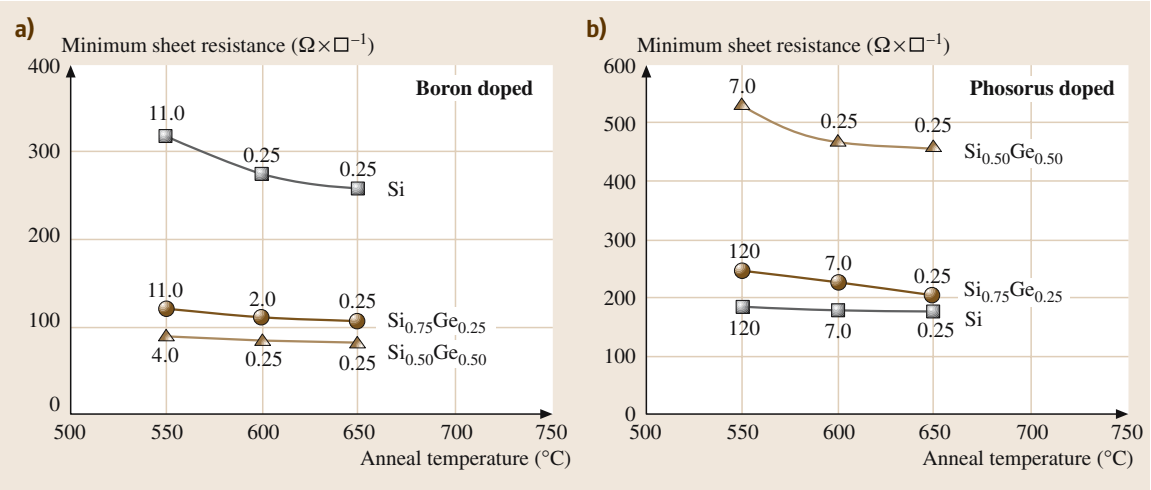


Fig. 22.30 Sheet resistance as a function of anneal temperature for boron- and phosphorus-doped polycrystalline Si_{1-x}Ge_x with various germanium contents (after Bang et al. [22.67], copyright 1995, American Institute of Physics)

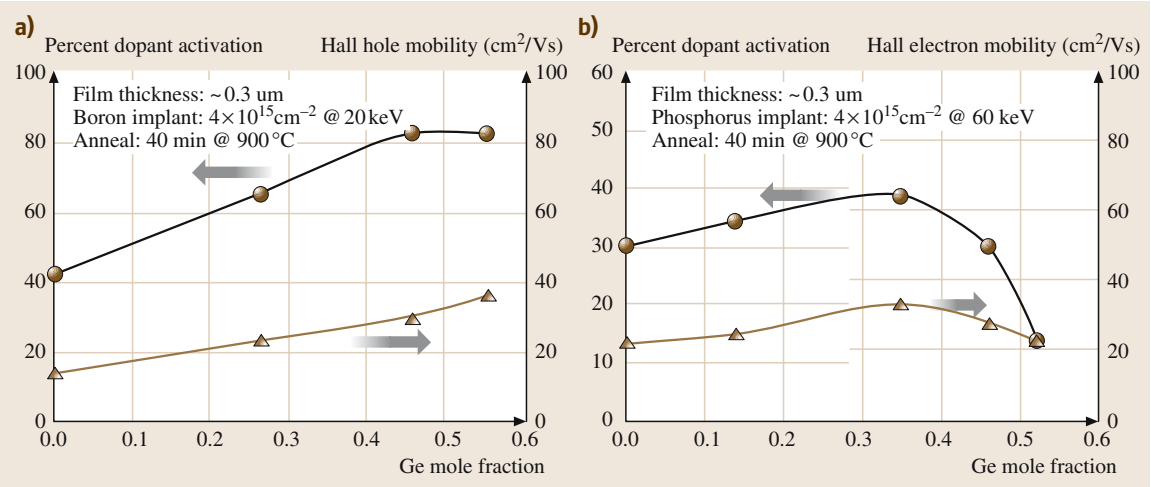


Fig. 22.31 Percentage dopant activation and Hall hole mobility as a function of germanium content for boron- and phosphorus-doped polySi_{1-x}Ge_x (after King et al. [22.60], copyright 1994, IEEE)

increases with increasing germanium content, with the increase being small between 0 and 25% germanium and large between 25 and 50%. Similar behaviour is seen for arsenic-doped polySi_{1-x}Ge_x where higher values of sheet resistance have been reported for polySi_{1-x}Ge_x than for polySi [22.64].

The explanation for the sheet-resistance results in Fig. 22.30 can be found in Fig. 22.31, which shows the results of Hall measurements [22.60]. For boron-doped polySi_{1-x}Ge_x both the activation and the Hall mobility increase with increasing germanium content, thereby explaining the decrease in sheet resistance with

increasing germanium content. For phosphorus-doped polySi_{1-x}Ge_x there is little change in activation and electron mobility at low germanium contents, but a sharp decrease in activation at germanium contents above 35%. This explains the sharp increase in sheet resistance seen in Fig. 22.30 for germanium contents between 25 and 50%. The decrease in activation at high germanium contents in the phosphorus-doped polySi_{1-x}Ge_x may be due to increased segregation at grain boundaries. Boron does not generally segregate to grain boundaries [22.68], which may explain the different behaviour in boron- and phosphorus-doped polySi_{1-x}Ge_x.

References

- 22.1 R. Braunstein, A. R. Moore, F. Herman: Phys. Rev. **109**, 695 (1958)
- 22.2 S. S. Iyer, G. L. Patton, J. M. C. Stork, B. S. Meyerson, D. L. Hareme: IEEE Trans. Electron. Dev. **36**, 2043 (1989)
- 22.3 C. A. King, J. L. Hoyt, J. F. Gibbons: IEEE Trans. Electron. Dev. **36**, 2093 (1989)
- 22.4 H. Miyata, T. Yamada, D. K. Ferry: Appl. Phys. Lett. **62**, 2661 (1993)
- 22.5 T. Vogelsang, K. R. Hofmann: Appl. Phys. Lett. **63**, 186 (1993)
- 22.6 J. Welser, J. L. Hoyt, J. F. Gibbons: IEEE Electron. Dev. Lett. **15**, 100 (1994)
- 22.7 A. Sadak, K. Ismile, M. A. Armstrong, D. A. Antoniadis, F. Stern: IEEE Trans. Electron. Dev. **43**, 1224 (1996)
- 22.8 B. Jagannathan, M. Khater, F. Pagette, J.-S. Rieh, D. Angell, H. Chen, J. Florkey, F. Golan, D. R. Greenberg, R. Groves, S. J. Jeng, J. Johnson, E. Mengistu, K. T. Schonenberger, C. M. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein, S. Subbanna: IEEE Electron Dev. Lett. **23**, 258 (2002)
- 22.9 Z. A. Shafi, P. Ashburn, G. J. Parker: IEEE J. Solid State Circuits **25**, 1268 (1990)
- 22.10 S. C. Jain, T. J. Gosling, J. R. Willis, R. Bullough, P. Balk: Solid State Electron. **35**, 1073 (1992)
- 22.11 J. M. Matthews, A. E. Blakeslee: J. Cryst. Growth **27**, 118 (1974)
- 22.12 J. M. Matthews, A. E. Blakeslee: J. Cryst. Growth **32**, 265 (1975)
- 22.13 R. People, J. C. Bean: Appl. Phys. Lett. **47**, 322 (1985)
- 22.14 S. Margalit, A. Bar-lev, A. B. Kuper, H. Aharoni, A. Neugroschel: J. Cryst. Growth **17**, 288 (1972)
- 22.15 O. W. Holland, C. W. White, D. Fathy: Appl. Phys. Lett. **51**, 520 (1987)
- 22.16 R. People: Phys. Rev. B **32**, 1405 (1985)
- 22.17 J. Poortmans, S. C. Jain, D. H. J. Totterdell, M. Caymax, J. F. Nijs, R. P. Mertens, R. Van Overstraeten: Solid State Electron. **36**, 1763 (1993)
- 22.18 T. Manku, A. Nathan: J. Appl. Phys. **69**, 8414 (1991)
- 22.19 T. Manku, A. Nathan: Phys. Rev. B **43**, 12634 (1991)
- 22.20 J. Poortmans: Low temperature epitaxial growth of silicon and strained Si_{1-x}Ge_x layers and their application in bipolar transistors; PhD thesis, University of Leuven (1993)
- 22.21 J. M. McGregor, T. Manku, A. Nathan: *Measured in-plane hole drift mobility and Hall mobility in heavily doped, strained p-type Si_{1-x}Ge_x* (Boston 1992) presented at Electronic Materials Conference
- 22.22 M. V. Fischetti, S. E. Laux: J. Appl. Phys. **80**, 2234 (1996)
- 22.23 J. Welser, J. L. Hoyt, J. F. Gibbons: IEEE Electron. Dev. Lett. **15**, 100 (1994)
- 22.24 C. W. Leitz, M. T. Currie, M. L. Lee, Z.-Y. Cheng, D. A. Antoniadis, E. A. Fitzgerald: J. Appl. Phys. **92**, 3745 (2002)
- 22.25 J. H. Bahng, K. J. Kim, H. Ihm, J. Y. Kim, H. L. Park: J. Phys.: Condens. Matter **13**, 777 (2001)
- 22.26 D. J. Robbins, L. T. Canham, S. J. Barnett, A. D. Pitt, P. Calcott: J. Appl. Phys. **71**, 1407 (1992)
- 22.27 N. L. Rowell, J.-P. Noel, D. C. Houghton, A. Wang, D. D. Perovic: J. Vac. Sci. Technol. B **11**, 1101 (1993)
- 22.28 D. A. Grutzmacher, T. O. Sedgwick, G. A. Northrop, A. Zaslavsky, A. R. Powell, V. P. Kesan: J. Vac. Sci. Technol. B **11**, 1083 (1993)
- 22.29 J. Brunner, J. Nutzel, M. Gail, U. Menczgar, G. Abstreiter: J. Vac. Sci. Technol. B **11**, 1097 (1993)
- 22.30 K. Terashima, M. Tajima, T. Tatsumi: J. Vac. Sci. Technol. B **11**, 1089 (1993)
- 22.31 H. Prestling, T. Zinke, A. Splett, H. Kibbel, M. Jaros: Appl. Phys. Lett. **69**, 2376 (1996)
- 22.32 L. Masarotto, J. M. Hartmann, G. Bremond, G. Rolland, A. M. Papon, M. N. Semeria: J. Cryst. Growth **255**, 8 (2003)
- 22.33 J. S. Park, T. L. Lin, E. W. Jones, H. M. Del Castillo, S. D. Gunapall: Appl. Phys. Lett. **64**, 2370 (1994)
- 22.34 S. S. Murtaza, J. C. Cambell, J. C. Bean, L. J. Petico-las: IEEE Photon. Tech. Lett. **8**, 927 (1996)

- 22.35 D.J. Robbins, M.B. Stanaway, W.Y. Leong, R.T. Carline, N.T. Gordon: Appl. Phys. Lett. **66**, 1512 (1995)
- 22.36 A. Chin, T.Y. Chang: Lightwave Technol. **9**, 321 (1991)
- 22.37 R. People, J.C. Bean, C.G. Bethia, S.K. Spitz, L.J. Peticolas: Appl. Phys. Lett. **61**, 1122 (1992)
- 22.38 P. Kruck, M. Helm, T. Fromherz, G. Bauer, J.F. Nutzel, G. Abstreiter: Appl. Phys. Lett. **69**, 3372 (1996)
- 22.39 R.A. Soref, L. Friedman, G. Sun: Superlattices Microstruct. **23**, 427 (1998)
- 22.40 G. Sun, L. Friedman, R.A. Soref: Superlattices Microstruct. **22**, 3 (1998)
- 22.41 J. Weber, M.I. Alonso: Phys. Rev. B **40**, 5684 (1989)
- 22.42 H. Landolt, R. Bornstein: *Numerical data and functional relationships in science and technology*, Vol. 11/17a, ed. by O. Madelung (Springer, Berlin Heidelberg New York 1982)
- 22.43 G.S. Mitchard, T.C. McGill: Phys. Rev. B **25**, 5351 (1982)
- 22.44 M. Meuris, S. Verhaverbeke, P.W. Mertens, M.M. Heyns, L. Hellemans, Y. Bruynseraede, A. Philipessian: Jpn. J. Appl. Phys. **31**, L1514 (1992)
- 22.45 A. Ishizaki, Y. Shiraki: J. Electrochem. Soc. **129**, 666 (1986)
- 22.46 B.S. Meyerson, F.J. Himpsel, K.J. Uram: Appl. Phys. Lett. **57**, 1034 (1990)
- 22.47 G.S. Higashi, Y.T. Chabal, G.W. Trucks, K. Raghavachari: Appl. Phys. Lett. **56**, 656 (1990)
- 22.48 F.W. Smith, G. Ghidini: J. Electrochem. Soc. **129**, 1300 (1982)
- 22.49 G. Ghidini, F.W. Smith: J. Electrochem. Soc. **131**, 2924 (1984)
- 22.50 J.L. Regolini, D. Bensahel, E. Scheid, J. Mercier: Appl. Phys. Lett. **54**, 658 (1989)
- 22.51 G.R. Srinivasan, B.S. Meyerson: J. Electrochem. Soc. **134**, 1518 (1987)
- 22.52 M. Racanelli, D.W. Greve, M.K. Hatalis, L.J. van Yzendoorn: J. Electrochem. Soc. **138**, 3783 (1991)
- 22.53 D.J. Robbins, J.L. Glasper, A.G. Cullis, W.Y. Leong: J. Appl. Phys. **69**, 3729 (1991)
- 22.54 M. Racanelli, D.W. Greve: Appl. Phys. Lett. **56**, 2524 (1990)
- 22.55 A. Ishitani, H. Kitajima, N. Endo, N. Kasai: Jpn. J. Appl. Phys. **28**, 841 (1989)
- 22.56 Y. Zhong, M.C. Ozturk, D.T. Grider, J.J. Wortman, M.A. Littlejohn: Appl. Phys. Lett. **57**, 2092 (1990)
- 22.57 Y. Kiyota, T. Udo, T. Hashimoto, A. Kodama, H. Shimamoto, R. Hayami, E. Ohue, K. Washio: IEEE Trans. Electron. Dev. **49**, 739 (2002)
- 22.58 J.M. Bonar: "Process development and characterisation of silicon and silicon-germanium grown in a novel single-wafer LPCVD system"; *PhD thesis*, University of Southampton (1996)
- 22.59 T.-J. King, K.C. Saraswat: IEDM Tech. Dig., 567 (1991)
- 22.60 T.-J. King, K.C. Saraswat: IEEE Trans. Electron. Dev. **41**, 1581 (1994)
- 22.61 J.A. Tsai, A.J. Tang, T. Noguchi, R. Reif: J. Electrochem. Soc. **142**, 3220 (1995)
- 22.62 V.D. Kunz, C.H. de Groot, S. Hall, P. Ashburn: IEEE Trans. Electron. Dev. **50**, 1480 (2003)
- 22.63 T.-J. King, J.R. Pfister, K.C. Saraswat: IEEE Electron. Dev. Lett. **12**, 533 (1991)
- 22.64 C. Salm, D.T. van Veen, D.J. Gravesteijn, J. Holleman, P.H. Woerlee: J. Electrochem. Soc. **144**, 3665 (1997)
- 22.65 Y.V. Ponomarev, P.A. Stolk, C.J.J. Dachs, A.H. Montree: IEEE Trans. Electron. Dev. **47**, 1507 (2000)
- 22.66 P. Ashburn: *Silicon-germanium heterojunction bipolar transistors* (Wiley, Chichester 2003)
- 22.67 D.S. Bang, M. Cao, A. Wang, K.C. Saraswat, T.-J. King: Appl. Phys. Lett. **66**, 195 (1995)
- 22.68 I.R.C. Post, P. Ashburn: IEEE Trans. Electron. Dev. **38**, 2442 (1991)

Gallium Arsenide

23. Gallium Arsenide

The history of gallium arsenide is complicated because the technology required to produce GaAs devices has been fraught with problems associated with the material itself and with difficulties in its fabrication. Thus, for many years, GaAs was labelled as “the semiconductor of the future, and it will always be that way.” Recently, however, advances in compact-disc (CD) technology, fibre-optic communications and mobile telephony have boosted investment in GaAs research and development. Consequently, there have been advances in materials and fabrication technology and, as a result, GaAs devices now enjoy stable niche markets.

The specialised uses for GaAs in high-frequency and optoelectronic applications result from the physical processes of electron motion that allow high-speed and efficient light emission to take place. In this review, these advanced devices are shown to result from the physical properties of GaAs as a semiconducting material, the controlled growth of GaAs and its alloys and the subsequent fabrication into devices.

Extensive use is made of chapters from “Properties of Gallium Arsenide, 3rd edition” which I edited with the help of *Prof. G. E. Stillman* [23.1]. This book was written to reflect virtually all aspects of GaAs and its devices within a readable text. I believe that we succeeded in that aim and I make no apologies in referring to it. Readers who need specialised data, but not necessarily within an explanatory text, should refer to the Landolt-Börnstein, group III (condensed matter) data collection [23.2, 3]. The sub-volumes A1 α (lattice properties) and A2 α (impurities and defects) within volume 41 are rich sources of data for all III–V compounds. Although there are no better sources than the original research papers, I have referred to textbooks where possible. This is because the presentation and discussion of scientific data is often clearer than in the original text, and these books are more accessible to students.

Gallium arsenide (GaAs) is one of the most useful of the III–V semiconductors. In this chapter, the properties of GaAs are described and the ways in which these are exploited in devices are explained. The limitations of this material are presented in terms of both its physical and its electronic properties.

23.1	Bulk Growth of GaAs	502
23.1.1	Doping Considerations	502
23.1.2	Horizontal Bridgman and Horizontal Gradient Freeze Techniques	503
23.1.3	Liquid-Encapsulated Czochralski (LEC) Technique	504
23.1.4	Vertical Gradient Freeze (VGF) Technique	506
23.2	Epitaxial Growth of GaAs	507
23.2.1	Liquid-Phase Epitaxy (LPE)	507
23.2.2	Vapour-Phase Epitaxy (VPE) Technologies	508
23.2.3	Molecular-Beam Epitaxy (MBE) ...	509
23.2.4	Growth of Epitaxial and Pseudomorphic Structures ...	511
23.3	Diffusion in Gallium Arsenide	511
23.3.1	Shallow Acceptors	512
23.3.2	Shallow Donors	513
23.3.3	Transition Metals	513
23.4	Ion Implantation into GaAs	513
23.5	Crystalline Defects in GaAs	514
23.5.1	Defects in Melt-Grown GaAs	514
23.5.2	Epitaxial GaAs (not Low Temperature MBE GaAs)	516
23.5.3	LTMBE GaAs	517
23.6	Impurity and Defect Analysis of GaAs (Chemical)	517
23.7	Impurity and Defect Analysis of GaAs (Electrical)	518
23.7.1	Introduction to Electrical Analysis of Defects in GaAs	518
23.8	Impurity and Defect Analysis of GaAs (Optical)	521
23.8.1	Optical Analysis of Defects in GaAs	521

23.9 Assessment of Complex Heterostructures	522	23.11.3 The High-Electron-Mobility Transistor (HEMT) or Modulation Doped FET (MODFET)	526
23.9.1 Carrier Concentration Measurements in Heterostructures.....	522	23.11.4 The Heterojunction Bipolar Transistor (HBT)	526
23.9.2 Layer Thickness and Composition Measurements	522	23.12 Devices based on GaAs (Electro-optical) .	527
23.10 Electrical Contacts to GaAs	524	23.12.1 GaAs Emitters	527
23.10.1 Ohmic Contacts	524	23.12.2 GaAs Modulators	531
23.10.2 Schottky Contacts	524	23.12.3 GaAs Photodetectors.....	531
23.11 Devices Based on GaAs (Microwave)	524	23.13 Other Uses for GaAs	532
23.11.1 The Gunn Diode	524	23.14 Conclusions	532
23.11.2 The Metal-Semiconductor Field-Effect Transistor (MESFET) ...	525	References	533

GaAs is one of the compound semiconductors that occur in the zincblende structure. As can be seen in Fig. 23.1, each Ga atom is tetrahedrally bonded to four As atoms in a structure that is similar to the diamond lattice but with alternating Ga and As atoms.

The presence of two types of atoms introduces a small component of ionic bonding into the structure and increases the bond strength above that of the group IV element (Ge) which Ga and As neighbour. As a result the band-gap energy E_g , is increased to 1.518 eV at $T = 0$ K ($E_g = 1.41$ eV at room temperature).

The lattice constant of GaAs at 300 K is 0.565 36 nm, giving a resultant density of 5.3165 g/cm³. At atmospheric pressure GaAs melts near 1238 °C but with the loss of As vapour.

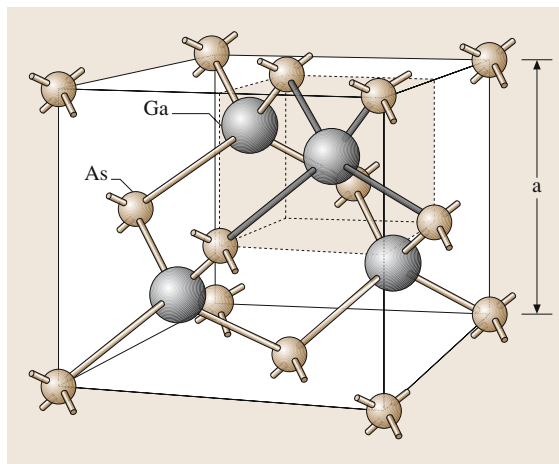


Fig. 23.1 The zincblende lattice of GaAs. The lattice constant a is indicated. (After [23.4])

Although all $\langle 001 \rangle$ atomic planes are equivalent this is not true for $\langle 011 \rangle$ and $\langle 111 \rangle$ planes. In particular, there are two types of $\langle 111 \rangle$ atomic planes; one terminated with Ga atoms only (the $\langle 111 \rangle_A$ planes), the other being terminated with As atoms only (the $\langle 111 \rangle_B$ planes). This makes the $\langle 111 \rangle$ directions strongly polar. A further consequence of the reduced symmetry of the lattice compared with the group IV semiconductors is that GaAs exhibits strong piezoelectric effects.

The easily cleaved directions are the $\langle 110 \rangle$ and $\langle 111 \rangle$ types. The $\langle 110 \rangle$ are the more useful as they allow square sections to be cleaved from a $\langle 001 \rangle$ section wafer, one of the major reasons for the use of $\langle 001 \rangle$ wafers in GaAs device fabrication.

The original interest in GaAs arose from its unusual band structure, compared to that of Si in Fig. 23.2, which demonstrates several attributes that are expected to be interesting for high-speed electronics and optoelectronics.

1. The fundamental band-gap energy of 1.41 eV (300 K) corresponds to a point in the diagram, the Γ point of the Brillouin zone, where both the crystal momentum for holes and electrons is zero. It follows that low-energy electrons and holes can recombine without the moderating influence of phonons that would otherwise be required to conserve momentum. Such direct recombination is expected to result in efficient emission of photons and GaAs would then be the base material for efficient light-emitting diodes (LEDs) and lasers.
2. The effective mass of conduction-band electrons is inversely proportional to the curvature at the bottom of the band and this curvature is considerably greater, and the effective mass is considerably smaller, in GaAs (about $0.063 m_e$) than in silicon (about $1.1 m_e$).

Table 23.1 Selected important properties of pure GaAs at 300 K

Lattice constant (Å)	5.653
Density (g/cm ³)	5.318
Band-gap energy (eV)	1.424
Band-gap type	Direct
Electron mobility (cm ² /Vs)	8500
Hole mobility (cm ² /Vs)	400
Zero-frequency dielectric constant	13.18
Conduction-band effective density of states (cm ⁻³)	4.45×10^{17}
Valence-band effective density of states (cm ⁻³)	7.72×10^{18}
Intrinsic carrier concentration (cm ⁻³)	1.84×10^6

Because the mobility of carriers is inversely proportional to their effective mass, it follows that free electrons in GaAs should be very mobile under the influence of an electric field, resulting in fast devices.

- An additional feature of Fig. 23.2 is the occurrence of subsidiary conduction-band minima near the six points of maximum momentum along the $\langle 100 \rangle$ directions. These are at the so-called L points of the Brillouin zone. These minima are degenerate and at an energy of 0.3 eV above the minimum energy of the conduction band at the Γ point. The relatively small curvature of these X minima results in a large effective electron mass of about $1.2 m_e$. Electrons at these points have a correspondingly low mobility. It follows that the excitation of low-effective-mass free electrons from the Γ point to the L minima, where they have high effective mass, can result in a *reduction* of electron velocity and hence electron current. This is the *transferred electron effect*. The current/voltage characteristics of such a device exhibit a negative differential resistance region, the Gunn effect, and this can be used to generate microwave radiation.
- One property of GaAs, not resulting from Fig. 23.2, is that the material can be produced in a high electrical resistivity, semi-insulating (SI) state. This allows devices to be fabricated on a near-insulating substrate. This is advantageous for high-frequency devices as the parasitic capacitances that occur between Si devices and their substrate, reducing their

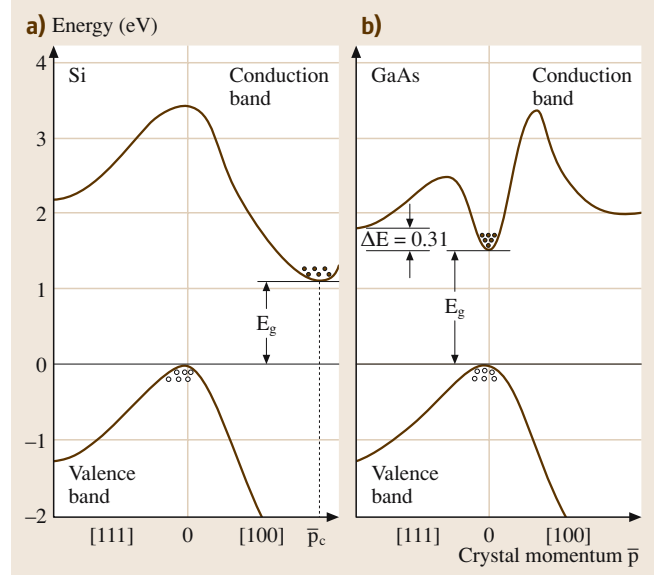


Fig. 23.2a,b The band structures of (a) Si and (b) GaAs. The primary band gap of Si is indirect with an energy E_g , of 1.12 eV. However, that of GaAs is direct (both free electrons and holes of low energy have states that correspond to zero crystal momentum at the Γ point of the Brillouin zone). Secondary conduction-band minima occur at 0.31 eV above the primary minimum near points of maximum crystal momenta along the $[111]$ -directions. These are at the L points. (After [23.4])

maximum operating frequency, will be absent in GaAs devices fabricated on SI substrates.

- A further property of great interest is that the lattice constants of GaAs and aluminium arsenide, AlAs, and their alloys, written as $\text{Ga}_{1-x}\text{Al}_x\text{As}$, where x is the atomic fraction of Al, are very similar. This allows different alloys to be grown consecutively without incurring defects due to lattice mismatch. Such heterostructures are not possible with Si and its known alloys.

The predictions made above will be addressed in later sections but first we consider the growth of GaAs single crystals suitable for device use. Subsequently, we consider the epitaxial growth of thin, GaAs-based layers, which are the basis of modern devices, means of processing and the fabrication and properties of important devices.

23.1 Bulk Growth of GaAs

In the simplest terms, GaAs crystals are synthesised by reacting together high-purity Ga and As. Dopants may be added to produce materials of different conduction type and carrier concentration. However, the growth of compounds from the melt is a thermodynamic process and is characterised by the shape of the phase diagram near the solidus at temperatures just below the melting point of the compound. Hurle has published a thorough reassessment of the thermodynamic parameters for GaAs, with and without doping [23.6]. The complete phase diagram for Ga–As, showing the GaAs compound, is shown in Fig. 23.3. However, details near the GaAs solidus are not revealed at this large scale.

The shape of the GaAs solidus is more complicated than indicated in Fig. 23.3. In particular, the congruent point, the maximum melting temperature, corresponds to a GaAs compound containing more As than exists in stoichiometric material (Fig. 23.4).

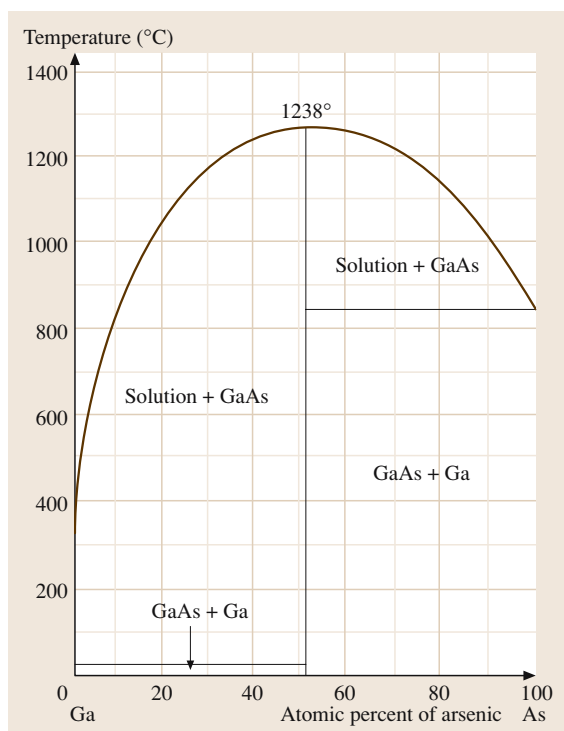


Fig. 23.3 The complete phase diagram of the Ga–As system at an assumed pressure of As vapour of about 1 atm. The vertical line near 50% composition is crystalline GaAs. (After [23.5])

It follows that growth from a stoichiometric melt results in the growth of a solid phase containing an excess of As of approximately $1 \times 10^{19} \text{ cm}^{-3}$. The solidus is retrograde so that, if thermodynamic equilibrium is maintained, this excess As must be lost from the host as the solid cools. If this As supersaturation cannot be relieved by out-diffusion from the crystal, then it will result in the generation of As-rich second phases (precipitates). Growth from a Ga-rich melt will result in the growth of a stoichiometric crystal. However, the melt will become more Ga-rich as growth proceeds, leading to the growth of solid GaAs that steadily becomes richer in Ga.

Experimental work has shown that the electrical properties of Ga-rich GaAs are poor (undoped Ga-rich GaAs is strongly p-type [23.7]) and this growth regime is rarely used. Although As richness is found in melt-grown GaAs, measured concentrations of excess As do not exceed 10^{17} cm^{-3} [23.8,9]. It is not known how most of the grown-in excess As atoms are lost.

23.1.1 Doping Considerations

Unlike group IV semiconductors where shallow donors are group V atoms only, donors in III–V compounds can occur on either sublattice, as IV_{III} or VI_{V} , where the

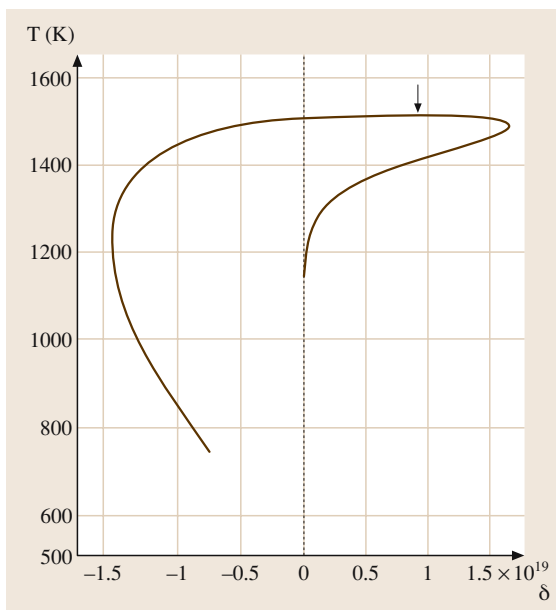


Fig. 23.4 The GaAs solidus. The arrow marks the congruent point (After [23.6])

capital letters represent the group in the periodic table from which the donor is selected and the subscript is the sublattice occupied. However, only certain group IV atoms have a suitable solubility on the Ga sublattice for them to be useful and the same situation is true for group VI atoms on As sites.

Similarly, acceptors can be of II_{III} or IV_{V} types. (In principle, group IV atoms can be both acceptors and donors, IV_{III} and IV_{V} respectively, a situation known as *amphotericity*. Silicon, and to a similar extent germanium atoms in GaAs, behave like this.) Once again not all of these occur in practice. The choice of dopant depends mainly on its segregation coefficient and solubility (see [23.6] for a discussion of this point). Figure 23.5 shows most of the known shallow donors and acceptors in GaAs with their ionisation energies. Also shown in this figure are two important impurities that generate levels deeper in the band gap. Cu is an unwanted deep acceptor contaminant but Cr has some technological importance and this element is discussed later.

In melt-grown GaAs, donors are usually either Si_{Ga} or Te_{As} , as each has a satisfactory segregation coefficient and solubility, although other group VI species have been used. Donors experience some degree of auto-compensation with the result that, up to a carrier concentration n of about $3 \times 10^{18} \text{ cm}^{-3}$, the carrier concentration is only about 80% of the added dopant concentration. Thermodynamic analysis has demonstrated that for group VI donors this is a result of the formation of (donor– V_{Ga}) pair defects. Auto-compensation for the case of Si doping is a result of the amphotericity of Si with a fraction, about 10%, of the Si atoms taking up As lattice sites and behaving as shallow acceptors. Above $n = 3 \times 10^{18} \text{ cm}^{-3}$, the carrier concentration tends to saturate and extra donor atoms are involved in vacancy complexes with a drop in electron mobility, the appearance of dislocation loops and a rapid rise in the lattice constant [23.6].

The p-type doping of melt-grown GaAs appears to be much easier. There is no auto-compensation and the solubility of Zn, the most common acceptor dopant in melt-grown GaAs, is over 10^{20} cm^{-3} . The segregation coefficient of Zn in GaAs is close to unity.

Although the partial pressure of Ga above a Ga–As melt at all reasonable temperatures is extremely small and can be ignored, the partial pressure of As vapour (As_2 and As_4) depends sensitively on the melt composition and this can be used to control the ratio of Ga/As in the melt as the crystal grows. Over a stoichiometric GaAs melt at the melting temperature of 1238°C this pressure is approximately two atmospheres (1 atm above

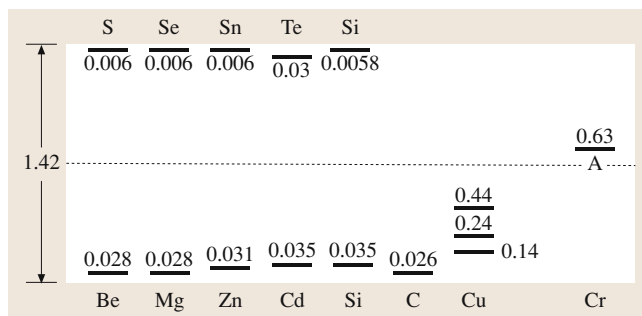


Fig. 23.5 The ionisation energies in eV of the shallow donors and acceptors in GaAs. Also shown are the important deep levels due to Cu and Cr. Cu is an impurity that can be introduced at growth or by processing and causes damaging deleterious effects in devices. Cr was originally used as a deep acceptor to produce semi-insulating GaAs. Its use has fallen since the introduction of LEC and similar growth techniques, see later. (Updated after [23.4])

ambient). It is necessary to control this pressure to retain melt composition and to impede the loss of As from the melt. This has resulted in the development of several growth techniques.

23.1.2 Horizontal Bridgman and Horizontal Gradient Freeze Techniques

The first growth method to be used commercially was the horizontal Bridgman (HB) technique [23.10]. In essence, the growth chamber consists of a horizontal quartz tube, at one end of which is a boat containing the Ga–As melt near the melting temperature (T_m) of 1238°C while the other end is cooler, near 617°C , and contains a small quantity of elemental As, Fig. 23.6.

This As produces the overpressure of As_4 that controls the stoichiometry of the Ga–As melt. The Ga–As melt is often formed by reacting metallic Ga in the growth section with As vapour produced from the As reservoir at the other end. In other processes polycrystalline GaAs is used as the source for the molten Ga–As. The growth can proceed in two ways. In the classical HB technique the initial melt exists in a region at a temperature just above T_m but adjacent to a small temperature gradient. A seed crystal, already placed in the crucible with part of it in the melt while the rest is at a temperature below T_m , will initiate growth. The melt is moved mechanically through the temperature gradient so that the entire melt is gradually cooled below T_m . Eventually, all of the GaAs melt solidifies on the seed as a single crystal and with the same orientation as the seed.

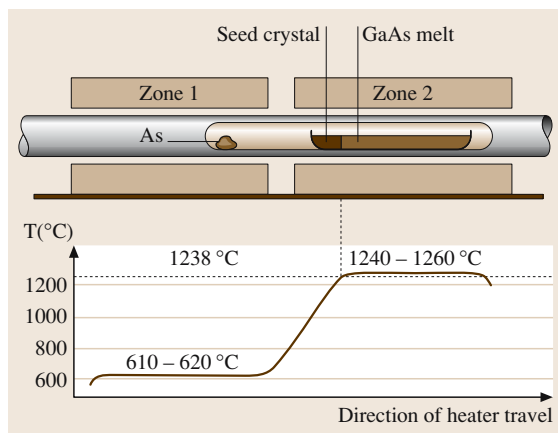


Fig. 23.6 The horizontal Bridgman technique for growing GaAs. The growth takes place in the high-temperature region of the furnace whilst the As vapour pressure is controlled by the presence of elemental As in the low-temperature region. The interface between the melt and the solid is gradually moved until all the GaAs is solidified. (After [23.4])

In a modification to **HB**, the mechanical withdrawal of the GaAs melt through the gradient is replaced by slowly reducing the temperature of the growth region electronically. This is the horizontal gradient freeze (**GF**) method [23.11]. It has the advantage of needing less room and is less sensitive to mechanical disturbance.

In both cases, crystals are best grown along $\langle 111 \rangle$ to reduce the occurrence of twinning. $\langle 001 \rangle$ wafers can be extracted from the $\langle 111 \rangle$ boule by accurate sawing. Advantages of the **HB** and horizontal **GF** methods include the good visibility of the growth procedure, allowing the operator to make modifications as growth takes place. However, grown crystals have a D-shaped cross section because of the shape of the melt in the boat and subsequent wafering. Considerable loss of material is incurred if the wafers are edge-ground to make them circular.

HB and horizontal **GF** GaAs are contaminated with silicon atoms from the quartz growth tube and this renders them n-type. This is not a problem if highly doped n-type material is required, as extra Si will be added to the melt anyway. Although there are few uses for p-type bulk GaAs, the over-doping of the melt with an acceptor species (usually Zn) effectively renders the crystal p-type. The reduction in hole mobility by compensation of a minority of the Zn atoms by the Si atoms is of little consequence.

However, the advantages of **SI** behaviour can only be obtained after the Si donors have been counter-doped

by the incorporation of Cr atoms in the melt [23.12, 13]. Chromium atoms act as deep acceptor centers in n-type GaAs and act to pin the Fermi energy just above the mid-gap, see Fig. 23.5. The GaAs so produced exhibits a resistivity that can exceed $10^8 \Omega\text{cm}$. Such Cr-doped GaAs was the mainstay of the high-speed GaAs device industry for over a decade, but growth problems, from the low segregation coefficient and solubility of Cr, resulted in low wafer yields, making the substrates expensive. Device fabrication problems from the rapid out-diffusion of Cr from the substrates into devices, resulting in lower speed, unacceptable device characteristics and device instability, also became evident.

Finally, the combination of poor thermal conductivity and low critically resolved shear stress (CRSS) at elevated temperatures results in inevitable thermal gradients in the cooling crystal producing plastic deformation by the creation of slip dislocations. In **HB** and horizontal **GF** material the dislocation densities are around 10^3cm^{-2} [23.14], although these densities can be reduced by over an order of magnitude due to *impurity hardening* in highly n-type GaAs [23.15]. Dislocations, their generation and their properties will be considered later.

23.1.3 Liquid-Encapsulated Czochralski (LEC) Technique

General Considerations

The Czochralski technique, in which a crystal is pulled from a melt, was originally used for metals and was then modified for the commercial growth of Ge and then Si. The rotation of the seed and crystal results in the immediate advantage of producing boules of circular cross section. Simultaneous rotation of the melt can result in efficient mixing of the host and dopant, a great advantage for doped crystals. In the case of compounds, where the partial pressures of one or both components are large, the melt surface must be protected either by incorporating an independent source of the vapour under dissociation (as in **HB** or **GF** growth) or by other means. In **LEC** growth, the surface is covered by an encapsulant layer of liquefied boric oxide, B_2O_3 [23.16, 17]. The axially rotating crystal is withdrawn through the encapsulant and then cools naturally by heat radiation, and by conduction and convection via the high-pressure ambient gas, see Fig. 23.7.

LEC pullers are high-pressure growth machines (and, therefore, expensive) because the internal pressure can exceed several atmospheres, especially when the Ga and As are compounded. In the growth of a compound

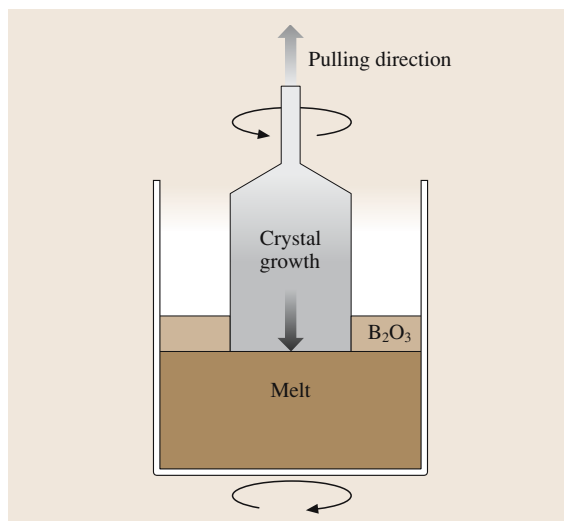


Fig. 23.7 LEC growth. The dark-grey arrow indicates the liquid–solid interface where growth takes place. As the growing crystal is pulled through the molten boric oxide it is rotated, resulting in a cylindrical boule. The boat is rotated in the opposite direction to stir the melt efficiently. (Courtesy Wafer Technology plc. UK)

like GaP, the internal pressures can exceed 100 atm. The boric oxide liquefies at temperatures of a few hundred degrees Celsius and floats to the surface of the other materials in the crucible. At all times during the subsequent reaction and growth it protects the surface from As loss. It also acts to purify the melt, probably because of the oxidising behaviour of its water content [23.18]. If the crucible is quartz, the resulting, nominally undoped, crystal is n-type because of Si incorporation. However, if the crucible is made from pyrolytic boron nitride (pBN) the crystal is SI over its entire length even without the introduction of Cr into the melt [23.19] (Sect. 23.1.3).

Nearly all LEC GaAs is contaminated by boron at concentrations up to 10^{18} cm^{-3} [23.20]. In all SI material these B atoms take up Ga lattice sites and are electrically inactive [23.21]. In Si-doped GaAs some of the B atoms appear to be incorporated as an acceptor species [23.22].

The automated growth of near-cylindrical crystals of accurately defined diameter would make LEC the choice of growth if it were not for the high dislocation density found in all but the highest carrier concentration n-type GaAs (this exception being a result of impurity hardening). Uncontrolled cooling of the LEC crystal causes the outside to contract on to the core, creating slip dis-

locations at densities of 10^4 – 10^5 cm^{-2} [23.23]. This defect creation occurs at temperatures just below T_m where dislocation motion by slip is easy. In addition, high concentrations of native point defects, which allow dislocation motion by climb, are also present. As a result, the dislocations are able to polygonise into cells, a rearrangement that reduces their strain energies. The final arrangement of a dislocation cell structure is the situation that is usually seen when a wafer is assessed.

As in HB and horizontal GF growth, the crystal can be doped n-type or p-type. Si or Te are normally used for n-type doping while Zn is used for p-type material. As mentioned above, an unfortunate reaction between the melt and the boric oxide has been found to occur with Si doping, leading to a reduction in Si uptake and considerable boron acceptor contamination of the GaAs [23.24]. In many cases, the concentration of B in the final crystal is comparable to the Si concentration. Nevertheless, Si remains one of the preferred donors for LEC GaAs.

LEC growth is controlled by computer and crystals of mass up to 20 kg and diameter up to 200 mm are easily produced. Wafers of 150 mm diameter are routinely supplied to device manufacturers from LEC crystals: at the time of writing this article 200-mm-diameter wafers were being made available to manufacturers for assessment.

Growth of SI LEC GaAs

The growth of nominally undoped GaAs by LEC from a pBN crucible results in SI behaviour. Chemical analysis of this type of GaAs always finds a concentration of carbon that is higher than the total concentrations of all other electrically active impurities [23.21]. The high carbon concentrations are not too surprising because not only is carbon a possible impurity in both Ga and As but there are many components of the LEC puller, namely the heaters and much of the thermal insulation, which are also made of carbon. Much work has shown that the carbon is introduced to the Ga–As melt through the gas phase, probably as carbon monoxide and the control of the partial pressure of this gas can be used to control the uptake of carbon in the crystal [23.25]. Because carbon atoms take up As sites and act as shallow acceptors, the resulting crystal would be expected to be p-type.

Some of the carbon acceptors are compensated by residual concentrations of shallow donors such as silicon and sulphur. The compensation of the rest of the carbon acceptors is performed by a native deep donor species, EL2, which pins the Fermi Energy close to the mid-gap. However, the final resistivity of the GaAs depends mainly on the carbon concentration [23.26]. The

atomic identity of EL2 was a hot topic of research for many years and there still remain questions as to its identity [23.27]. However, what is not questioned is that the defect involves the As-antisite defect, As_{Ga} .

After growth, high concentrations of EL2 are found associated with dislocations and this results in nonuniformities in electrical properties [23.28, 29]. Micro-precipitates of hexagonal As are also found in close association with the dislocations [23.30]. Most manufacturers use ingot anneals to render EL2 concentrations more uniform and to improve electrical uniformity [23.31]. Some follow ingot heat treatments with anneals of the individual wafers [23.32]. The schedules of these treatments vary between wafer suppliers but, in general, after these treatments the resistivity is of the order of $10^7 \Omega \text{ cm}$ with a uniformity of better than $\pm 10\%$.

It must be emphasised that the dislocation density cannot be reduced by heat treatments and the needs of device manufacturers for material of lower dislocation density has led to the development of improved growth techniques, the one now in general use being the vertical gradient freeze (VGF) method.

23.1.4 Vertical Gradient Freeze (VGF) Technique

This is a modification of the horizontal technique where the melt is contained in a vertical crucible above a seed crystal. The crucible, surrounded by the furnace, is pBN if SI GaAs is to be grown; otherwise, it can be quartz. The growth proceeds from the bottom of the melt upwards until the melt is exhausted [23.33]. Because cooling is better controlled, the resulting temperature gradients are much reduced compared to the LEC method, resulting in crystals with lower dislocation densities, typically $10^2\text{--}10^3 \text{ cm}^{-2}$. A schematic of VGF method is shown in Fig. 23.8.

The starting material is polycrystalline GaAs. Often, this can be synthesised from the elements in a LEC puller before being withdrawn rapidly from the crucible. This is a useful way of employing a LEC puller that might otherwise be redundant. The LEC puller also allows the introduction of a controlled amount of carbon, which will be necessary for SI behaviour in the final ingot. The use of polycrystalline GaAs starting material means that the VGF equipment can be easily fabricated in-house, as

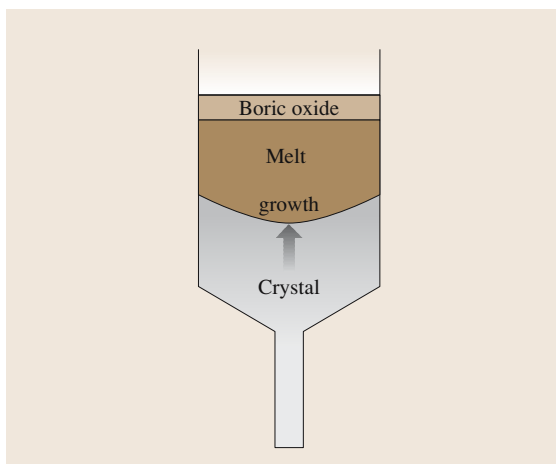


Fig. 23.8 VGF growth. The seed crystal, in the narrow section at the bottom of the crucible, is held below the melt and the growth takes place upwards. The *solid arrow* indicates the solid–liquid interface where growth takes place. The boric oxide surrounds the liquid and solid and helps in the final extraction of the cooled crystal from the crucible. (Courtesy Wafer Technology plc. UK)

high pressures will not be experienced. Boric oxide can be used to encapsulate the melt and to ease crystal removal from the crucible but the sealed system will ensure that the melt remains nearly stoichiometric [23.34].

Initially the major drawback to VGF was the low growth rate coupled with the inability to see the progress of the growth. In other words, if growth was not progressing correctly, this could not be detected until after the entire melt was solidified and this was several days in most circumstances. However, after the method is optimised the yield can approach 100%. The usual dopants can be employed if conducting material is required.

The reduced dislocation density in VGF GaAs has made it the material of choice for most applications and especially in optoelectronics, where dislocations are particularly deleterious to performance as they act as nonradiative recombination paths. At the time of writing, SI wafers of 200 mm diameter have been supplied as test wafers to device manufacturers.

Readers who require a recent and more thorough discussion of the growth of GaAs from the melt are referred to [23.35].

23.2 Epitaxial Growth of GaAs

Very few devices now use melt-grown GaAs in their active parts. Instead, they rely on the substrate to act as support for complex structures that are grown on the surface. These will typically employ ternary or even quaternary alloys. The former are the most important and are exemplified by the $\text{Ga}_{1-x}\text{Al}_x\text{As}$ alloys, all of which have a very similar lattice constant as GaAs, varying from 5.6533 Å for GaAs to 5.6605 Å for AlAs. Selected electrical and optical properties of these alloys are presented in Fig. 23.9.

These alloys are in the form of thin layers which must have an accurately controlled composition, lattice-match the substrate, have accurate doping and, most of all, a well-defined layer thickness, often of only a few hundred nm. Growth of a new material on to a substrate where the atomic planes accurately line up is *epitaxial*. In this section, we present four approaches to the epitaxial growth of GaAs, the first two being more of historical than practical interest.

23.2.1 Liquid-Phase Epitaxy (LPE)

In **LPE**, the GaAs layer to be deposited is formed by first dissolving GaAs and the dopant into a liquid, usually Ga, at elevated temperatures, normally around 800 °C [23.37]. The solubility of GaAs in Ga is 3×10^{-3} mass fraction at these temperatures but it falls rapidly if the temperature is reduced. This behaviour can be seen in Fig. 23.3. Thus, if the Ga melt is saturated at a high temperature and is placed over the substrate, GaAs will be deposited as the temperature is reduced. Dopants added to the Ga melt will also be introduced into the growing GaAs layer, allowing junctions to be fabricated. If the melt is replaced by a molten (Ga + Al) alloy, a layer of GaAlAs can be grown, also with doping.

LPE usually uses several pots of molten metal (Ga or Ga + Al, for example) in a high-purity graphite, slider system. These can be moved over the substrate in turn to allow the growth of consecutive layers. A figure of a sliding-boat **LPE** reactor is shown in Fig. 23.10.

Advantages of **LPE** include the high crystalline purity of layers especially in terms of low concentrations of native point defects. Disadvantages are associated with the control of growth. It is difficult to produce reliably flat layers as convection currents in the liquid metal tend to produce a rippled surface. The thickness often

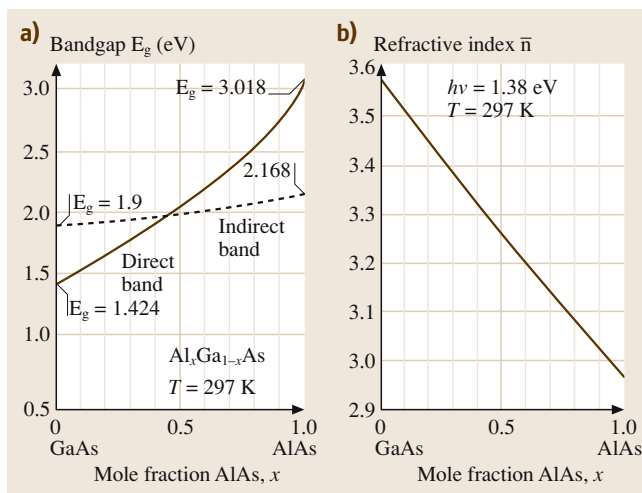


Fig. 23.9a,b Compositional dependence of some important properties of AlGaAs . **(a)** shows the dependence of the band-gap energy on composition. Below 45% Al, the alloy has a direct band-gap energy that increases steadily with Al composition. However, above 45% the band gap becomes indirect and the band-gap energy of pure AlAs is only 2.168 eV. **(b)** The refractive index of the AlGaAs alloys shows no dependence on the band-gap type and decreases steadily from a value of approx. 3.6 (GaAs) to 2.97 (AlAs), both these values being wavelength-dependent. (After [23.4])

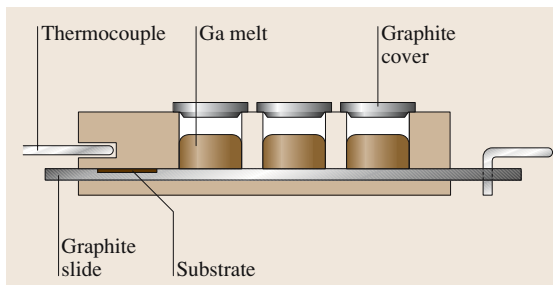


Fig. 23.10 **LPE** growth using the slider system. The Ga melts saturated with GaAs and dopants are sequentially placed over the GaAs substrate allowing several epitaxial layers to be grown. (After [23.36])

increases towards the periphery of the substrate and it is difficult to produce large areas of constant thickness. **LPE** was important for the **LED** and laser industry in early years but it is now rarely used for GaAs.

23.2.2 Vapour-Phase Epitaxy (VPE) Technologies

Chloride and Hydride Growth

In **VPE**, the components of the layer to be grown are transported to the substrate surface in the form of gases.

In chloride and hydride growth, the As vapours are AsCl_3 and AsH_3 , respectively, both of which are converted to gaseous As_4 in the growth reactor. The volatile Ga component is GaCl produced by passing HCl gas over a well containing heated Ga metal. This takes place near 850°C . The volatile Ga and As components are reacted with H_2 over the GaAs substrate, which is held near 750°C . A two-zone furnace is therefore required. The growth rate can be up to $20\ \mu\text{m}$ per hour. Dopants can be introduced as chlorides. Full details of these growth techniques have been reviewed by Somogyi [23.38].

The major commercial use for these growth methods is for the fabrication of $\text{GaAs}_{1-x}\text{P}_x$ LEDs, see Fig. 23.11.

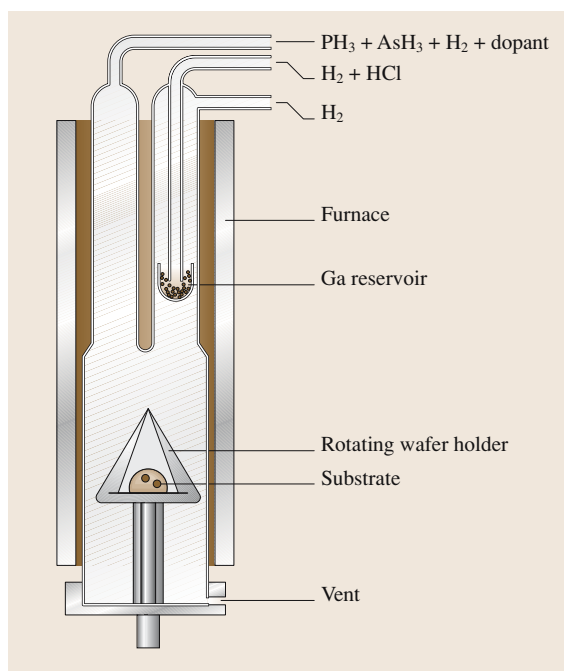


Fig. 23.11 The hydride growth method. A vertical system is shown but horizontal methods are also used. The volatile gallium chloride GaCl is produced near 850°C , and growth takes place near 740°C . The rotating sample holder ensures that the growth on several substrates is uniform and reproducible. (After [23.36])

Phosphorus can be introduced as PCl_3 or PH_3 and the growth is sufficiently fast that grading from pure GaAs to the required alloy can be performed in a reasonable time by varying the ratio of AsCl_3 to PCl_3 partial pressures. Unfortunately, modern device technologies rely heavily on alloys from the GaAs/AlAs system and growth of AlAs is not possible using these methods because no chloride of Al is sufficiently volatile.

Metalorganic Chemical Vapour Deposition (MOCVD)

In **MOCVD**, sometimes known as metalorganic vapour-phase epitaxy (**MOVPE**), the gaseous As precursor is AsH_3 and the volatile metal sources are alkyls (sometimes known as metalorganics) such as trimethyl-gallium (**TMG**), $\text{Ga}(\text{CH}_3)_3$, although others have been used. Trimethyl-aluminium (**TMA**), $\text{Al}(\text{CH}_3)_3$, is used for the Al component of the alloy and trimethyl-indium (**TMI**) is used for the In component.

For p-type material the usual acceptor is Zn in the form of diethyl zinc, $\text{Zn}(\text{CH}_3)_2$. Si is the most commonly used donor in the form of silane, SiH_4 .

All the alkyls are volatile liquids at room temperature and are kept refrigerated in order to reduce and control their vapour pressures. AsH_3 and SiH_4 are gases and are supplied diluted in hydrogen.

The alkyls are released from their cylinders by bubbling high-purity hydrogen through them. They are then introduced to all the other constituents with more hydrogen at room temperature. The reaction to produce the layer only occurs when the vapours are passed downstream over the heated substrate. This is held at approximately 750°C for GaAs growth but is somewhat higher for AlGaAs. Figure 23.12 is a schematic of a **MOCVD** apparatus.

This technique has been discussed in detail [23.39, 40]. Once the control of the gas flow over the heated substrate has been optimised, layers can be grown with great ease and with great accuracy, the layers being of high crystalline quality, and of good uniformity in both constitution and thickness. The growth of alloys of different composition or material with varying dopant concentration can be simply performed by changing the hydrogen flow rate through the alkyls or dopant sources using mass-flow controllers.

Note that SI epitaxial GaAs cannot be obtained unless Cr is added during growth or lightly p-type GaAs is grown and a post-growth anneal to generate EL2 centres is undertaken. Both methods have been reported, (see [23.41–44] for further details).

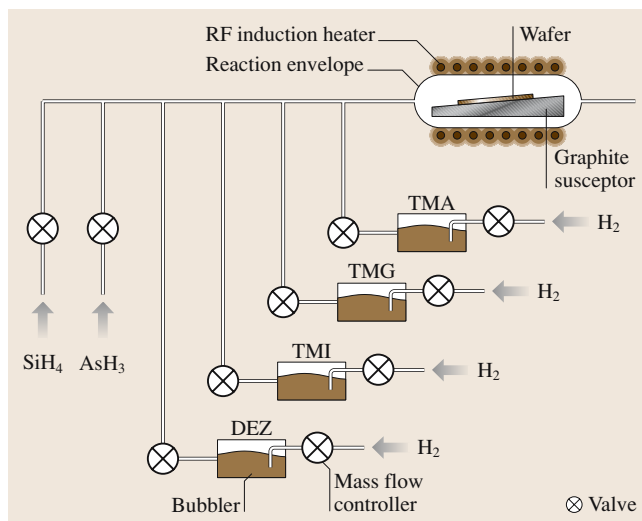
Fig. 23.12 MOCVD growth. Separate mass-flow controllers and valves control the hydrogen flow through each of the alkyl liquids, which are held in cooled, sealed flasks. The metalorganic vapours are introduced to the substrate with arsine. The substrate is heated by an induction furnace via a graphite susceptor that supports the substrate as well as coupling to the field of the RF field to produce heating. The gaseous silane is for n-type doping

In a research laboratory, the growth equipment can be built in-house and a skilled operator can control the growth of single samples. Commercially the equipment is much more complicated, with several substrates being processed together and the entire growth procedure being computer-controlled. Manufacturers put much effort into ensuring that the gas flow over each substrate is as uniform as possible to ensure reproducible growth. Early problems with difficulties in obtaining alkyls of satisfactory and reproducible quality have been solved.

However, a limitation of MOCVD concerns the rate at which a layer composition can be changed. The abruptness of the interface between layers can be critical in many modern applications and this is often limited in MOCVD by mixing of the gas upstream from the substrate. Thus, even if the mass-flow controllers that control the rate at which hydrogen bubbles through the alkyls change their relative concentrations instantly, there will be some mixing in the upstream gases which will cause this change to be more gradual when the gases flow over the substrate. This causes a slight grading in the grown interface and this can be deleterious to some devices. Nevertheless, most of current GaAs-based devices are produced commercially by MOCVD.

Unfortunately, all the precursors used for MOCVD are poisonous, with AsH_3 being extremely so. Moreover, satisfactory growth requires a surplus of AsH_3 over the alkyls of about ten to one in atomic concentrations to be present. It follows that the exhaust gases

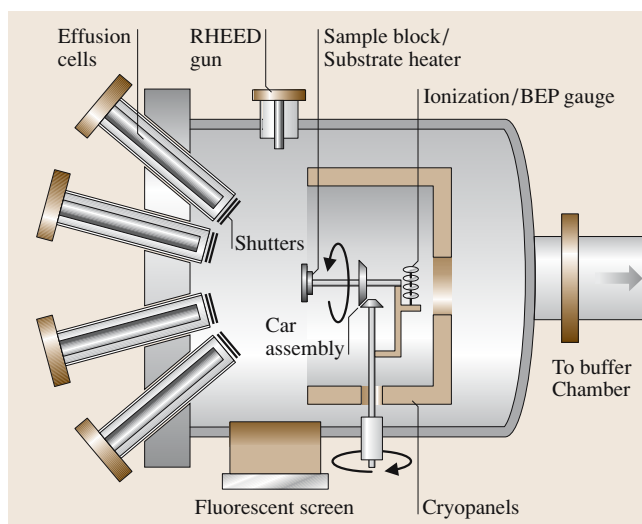
Fig. 23.13 MBE growth takes place in a UHV chamber. Each component of the required epitaxial layer is supplied by an effusion cell (also known as a Knudsen cell) and shutter. The substrate is held on a heated, rotating holder. Cryopanels cooled by liquid nitrogen help to keep the base pressure of the system to about 10^{-11} Torr. The RHEED gun and associated fluorescent screen allow the growth process to be monitored continuously. Substrates are positioned on the sample block via the buffer chamber where they have been previously prepared



are As-rich and very toxic. MOCVD equipment requires considerable exhaust gas cleaning and this makes the technique rather demanding in terms of effluent processing.

23.2.3 Molecular-Beam Epitaxy (MBE)

At its simplest, MBE is a high-purity growth technique where all the components of a required material are evaporated as elemental beams. The term molecular reflects the observation that some elements evaporate as molecules. As an example, under MBE growth conditions arsenic vapour is primarily in the form of As_4



molecules. The beams are incident on a heated, rotating substrate where they react to form a growing layer [23.45].

In practice, the growth takes place in an ultra-high-vacuum (UHV) chamber at a base pressure of about 10^{-11} Torr. The sources are usually high-purity elements held in Knudsen cells (high-vacuum effusion cells). The molecular beams are controlled by simple metal shutters that can be moved rapidly in and out of the beams by external drives. A typical MBE reactor will have up to eight such sources, some for the components of the host semiconductor, Ga, Al, etc. and others for dopants, each held at the relevant temperature to produce a molecular beam of the correct effective pressure, see Fig. 23.13.

The use of UHV technology allows the use of in situ analysis of the growth to be monitored and this has provided much information on details of the growth mechanisms. A favourite tool is reflection high-energy electron diffraction (RHEED) whose use gives important information regarding the quality of the growth with single atomic layer sensitivity.

The advantages of MBE include the use of high-purity elemental sources rather than compounds (whose purity is less controllable). The use of moveable shutters in front of each cell means that each source can be turned on or off in a fraction of a second. As a result, MBE-grown material can be of very high quality with the sharpest interface abruptness. These qualities are essential in several microwave devices and this is one of the major commercial uses for MBE. However, MBE has been a favourite growth technique for semiconductor research laboratories because it is well suited for the small-scale growth of specialised structures.

As an industrial technique, MBE suffers from several problems. The first results from its reliance on UHV technology, meaning that it is expensive to install, requires large quantities of liquid nitrogen to keep cold, considerable power to keep under vacuum, and special clean rooms. It is expensive to operate. Conditions for correct growth include a beam effective pressure ratio of As to group III metals of about six. As a result, the inside of the machine becomes coated with arsenic, making it very unpleasant to clean. Replenishing the cells when they are exhausted can be a slow and difficult process because recovery of the vacuum after the inside of the machine is exposed to atmospheric gases can take several days.

There are three modes for MBE growth: normal, low temperature and gas source.

Normal MBE Growth

In conventional MBE [23.46], the growth normally takes place at around 600°C with an As_4 flux that is considerable greater than the Ga flux. Often the beam effective pressure (BEP) ratio of As to Ga is around six. The Ga flux controls the growth rate and a Ga flux of $6 \times 10^{14} \text{ cm}^{-2} \text{ s}^{-1}$ produces a growth rate of one monolayer per second.

The GaAs so produced is of excellent structural and electrical quality. It is usually doped n-type with Si but Se and Te are often used. Unlike melt-grown GaAs, there is little auto-compensation and doping levels exceeding 10^{19} cm^{-3} can be obtained; p-type doping is normally achieved using Be, although for some devices carbon is preferable. These acceptors can be incorporated at concentrations exceeding 10^{20} cm^{-3} . AlGaAs alloys are grown by opening the shutter of an Al cell. Growth proceeds at a somewhat higher temperature but with similar ease. Should AlGaAs alloys of different composition be required, it is usual to have two Al cells at different temperatures and these are opened as required. For similar reasons a device structure with different n-type doping regions such as a metal–semiconductor field-effect transistor (MESFET) or high-electron-mobility transistor (HEMT) will require the use of at least two Si cells at different temperatures. These means of changing material composition increase equipment complexity and cost.

Low-Temperature MBE (LTMBE) Growth

Unlike the VPE techniques presented previously, MBE growth can proceed at temperatures as low as 200°C , in a regime called low-temperature MBE (LTMBE), see [23.47, 48]. One reason for this is that no chemical reactions are needed to release the elemental Ga and As on the GaAs surface from gaseous precursors (although the As_4 must be dissociated) and there is sufficient surface atomic mobility at these low temperatures to ensure uniform, crystalline growth. At these low growth temperatures and under standard As-rich growth conditions with a BEP ratio of greater than 3, undoped GaAs is very As-rich, with up to 1.5% excess As, contains high concentrations of point defects and is high resistivity. This material is heat-treated before use to improve its properties and stabilise it. Growth of LTMBE GaAs under less As-rich conditions, or at higher temperatures, results in nearly stoichiometric material. In all cases, however, SI behaviour is observed.

The real advantage of annealed LTMBE GaAs to device manufacturers is the extremely low minority carrier lifetime of $< 0.5 \text{ ps}$. This makes it useful in the fabrica-

tion of ultrafast optodetectors [23.49] and as a substrate for high-speed integrated circuits (ICs) [23.50].

Gas-Source and Metalorganic MBE

In modifications known as gas-source MBE (GSMBE) and metalorganic MBE (MOMBE), some sources that are volatile at room temperature are used.

In GSMBE, AsH₃ replaces the metallic As source and some dopants are also replaced by volatile sources. One of the advantages of GSMBE is the virtually unlimited source of As (from a high-pressure gas bottle) and the ease of replacing it and similar group V sources. These are exhausted rapidly in normal growth because of the need to use an excess of these elements.

In MOMBE the group III metallic sources are replaced by metalorganics like TMG. By using a combination of triethyl-gallium (TEG) and TMG, MOMBE growth can lead to the controlled incorporation of carbon acceptors at concentrations up to 10^{21} cm^{-3} , necessary for heterojunction bipolar transistors (HBT) fabrication (see later) and for other uses where stable, highly p-type GaAs is required. Similar ease of carbon incorporation in GSMBE uses CCl₄ or CBr₄ as sources. A thorough review of these techniques has been given by Abernathy [23.51].

23.2.4 Growth of Epitaxial and Pseudomorphic Structures

The epitaxial growth of GaAs on GaAs is one of the simplest epitaxial growth processes where, if the growth is performed correctly, analysis of the resultant structure will show no break in the atomic planes at the substrate-layer interface. Because the lattice constants of the ternary GaAlAs alloys match that of GaAs, the growths of these are also simple in principle and are the basis for many devices.

Attempts to grow layers with lattice constants that are dissimilar to the substrate result in the incorporation of misfit dislocations. Although these are initially parallel to the interface, they are deleterious to device

performance because their interactions can cause them to orient themselves in the growth direction and to invade the electrically active structure above. They can then act as recombination centres. However, even when a required ternary does not lattice-match the substrate two approaches can be used to give satisfactory device results; grading and pseudomorphic growth.

In grading, the material composition, and hence the lattice constant, is changed gradually from that of the substrate to that of the required alloy. Misfit dislocations are still produced but they are now distributed on many lattice planes and they tend not to interact. This restricts them to the graded region and away from the active regions above. Grading is used for the production of low-cost visible LEDs based on GaAs_{1-x}P_x. These devices can be formed on either GaP or GaAs substrates according to the required alloy composition (which controls the colour) and growth by the hydride or chloride process allows grading to be performed. Other uses for grading occur in the production of optical confinement layers in certain types of lasers (Sect. 23.13).

Theoretical calculations show that mismatched growth does not result in the immediate creation of misfit dislocations, because there exists a critical thickness below which the strain energy is insufficient to produce them [23.52]. If growth is stopped at this stage, the layer remains dislocation-free. Indeed, if a layer of the substrate compound now covers the layer, the structure is further stabilised. Its enforced lattice match to the substrate strains the mismatched layer; if it would normally be cubic, the epilayer assumes a tetragonal structure, with the *c*-axis aligned in the growth direction. This distortion produces electrical and optical properties that are no longer isotropic. In other words, properties parallel to the growth interface are different to those perpendicular to it. Often these offer advantages over the cubic material. This type of growth is called pseudomorphic and is very useful in several microwave devices. Pseudomorphic structures can be grown by MOCVD or MBE, although the latter is usually preferred.

23.3 Diffusion in Gallium Arsenide

Diffusion, which refers to the motion of impurity atoms under a concentration gradient, can be a positive or negative process. As an example of the first category, diffusion can be used to introduce impurity atoms into the bulk of a solid. This is the most important process in Si technology, for example. Diffusion is also

one of the atomic processes that are thought to be important in making effective Ohmic contacts. However, diffusion is also the mechanism by which large concentration gradients grown in to a structure can relax and cause degradation to the characteristics of the device.

Tuck has written excellent reviews of atomic diffusion in semiconductors [23.53, 54]. For various reasons, the diffusion characteristics of most impurities in GaAs are only poorly understood. In this section, we present a short summary of those diffusion parameters that are well understood or are important in GaAs.

Diffusion is a high-temperature process resulting from the random motion of the impurity atoms within the host matrix. The process is described by the two Fick's laws.

The first law states that the flux of diffusing atoms J in a given direction, say x , is proportional to the concentration gradient along x or $J = -D(\partial C/\partial x)$. The negative sign shows that the net motion of atoms is in the opposite direction to the concentration gradient. This law can easily be proved by atomic models. The constant of proportionality D is the diffusion coefficient and is a strong function of temperature T according to $D = D_0 \cdot \exp(-E/kT)$, where E is an activation energy for the process and D_0 is a pre-exponential function, the diffusion constant.

There are two simple diffusion mechanisms. The first is the *substitutional* mechanism where impurities on a sublattice (Ga or As) move by jumping into adjacent vacant lattice sites. E has two components which add: an energy to move from one site to another, the migration energy, E_m , and a further *energy of formation* to produce the vacancy, E_v . This results in a large value of E and slow diffusion, normally undetectable at diffusion temperatures under a few hundred degrees Celsius below the melting temperature. The second is the *interstitial* mechanism where impurity atoms in interstitial sites move by jumping into adjacent interstices. Here E has a single component, the migration energy, of a value close to E_m . This results in a small value of E and rapid diffusion. Interstitial diffusion can often be observed at low temperatures, in extreme cases at or below room temperature.

The second law represents conservation of mass and states that the rate at which a concentration changes in a given volume, $\partial C/\partial t$, is equal to the difference in input and output fluxes from the volume. Combination of the two laws, for a value of D that does not depend on concentration, results in the diffusion equation, $\partial C/\partial t = D(\partial^2 C/\partial x^2)$. Solutions to this equation have been calculated for many situations including those that are of experimental interest [23.56]. When experimental diffusion profiles do not match theoretical predictions, it must be concluded that D is a function of C .

Experimental investigations usually start with coating the surface of a GaAs sample with the impurity.

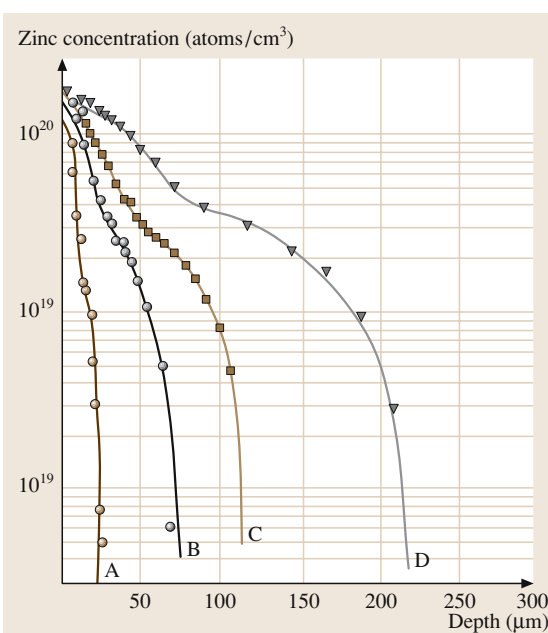


Fig. 23.14 Experimental diffusion profiles for Zn in GaAs at 1000 °C with excess arsenic in the ampoule. Diffusion times: A: 10 min, B: 90 min, C: 3 h, D: 9 h. (After [23.55])

The GaAs is heated rapidly to temperature T and held at T for a known time. The GaAs is then cooled rapidly. Concentrations of the impurity as a function of depth are then determined, either by using radiotracer methods (if the impurity can be tagged with a relatively long half-life isotope), by secondary-ion mass spectrometry (SIMS) if this technique is available, or electrically if the impurity is electrically active at room temperature.

Diffusions at different temperatures can be compared via an Arrhenius plot to derive the activation energy E . Diffusions from a thin layer are expected to result in a smooth, exponential profile. Other diffusion conditions result in profiles of different forms [23.53]. It is necessary to consider the loss of As from the GaAs at elevated temperatures when undertaking this procedure and a little elemental As is often added to the diffusion ampoule to replace loss from the surface. Other complications include possible chemical reactions between the impurity and either Ga or As, or both, which can result in severe changes to the experimental conditions. For these reasons, many diffusion measurements quoted in the literature may be quite inaccurate.

23.3.1 Shallow Acceptors

The first impurity whose diffusion profile in GaAs was carefully analysed was zinc because it was by diffusion of Zn into n-type GaAs to make a p–n junction that early LEDs were fabricated. When Zn is diffused into GaAs (Fig. 23.14) the profile is found to be deeper than expected for a substitutional mechanism. Moreover, it is not exponential but exhibits a concave shape with a very abrupt diffusion front.

Electrical evaluation reveals that, after diffusion, nearly all the Zn atoms act as shallow Zn_{Ga} acceptors. Interpretation suggests two possible *interstitial–substitutional* diffusion mechanisms, both of which involve the diffusion of singly positively charged interstitial Zn atoms, Zn_i^+ , which take up Ga lattice positions to become substitutional Zn_{Ga}^- acceptors. Either of these models suggests that D is proportional to $[\text{Zn}_{\text{Ga}}^-]^2$.

Under certain circumstances, the diffusion of Mn, which is a shallow acceptor, Mn_{Ga} , can be very similar to Zn. However, surface reactions cause experimental problems so that data is not so consistent.

Be_{Ga} is an acceptor often used in MBE growth. Its diffusion behaviour is poorly investigated but it has been concluded that D is concentration-dependent. Rapid ageing characteristics of GaAs/AlGaAs heterojunction bipolar transistors (HBTs) have been attributed to diffusion of Be from the highly p-type base into the emitter and collector regions [23.57].

C_{As} is a shallow acceptor and is important in Si GaAs as well as in MBE growth. It seems to be a very slow diffusing species, probably via purely substitutional diffusion. The diffusion rate depends on whether the GaAs is As- or Ga-rich, being greater in the former case. Arrhenius plots give expressions for D as $0.110 \exp(-3.2 \text{ eV}/kT)$ and $2.8 \times 10^{-4} \exp(-2.7 \text{ eV}/kT) \text{ cm}^2 \text{ s}^{-1}$, respectively. Carbon diffusion in GaAs has been reviewed by Stockman [23.58].

23.3.2 Shallow Donors

All shallow donors of scientific interest, group IV atoms occupying Ga or group VI atoms occupying As lattice

sites, are slow substitutional diffusers whose measurement has posed considerable problems [23.54]. Donor diffusion appears neither to have any direct technical application nor to be involved in any device degradation, although it is expected to be involved with the creation of good Ohmic contacts to n-type GaAs. A recent assessment of Si diffusion in GaAs using SIMS to trace the motion of the Si atoms has been reported in [23.59].

In general, it is observed that atoms occupying the group V sublattice are generally very slow diffusing, possibly a result of the high energy of formation of V_{As} .

23.3.3 Transition Metals

Transition metals normally occupy Ga lattice sites and act as deep acceptors. Cronin and Haisty [23.12] grew GaAs crystals doped with virtually all the first series of transition metals and were the first to observe their acceptor properties. This investigation showed that Cr doping produced SI material. However, most of these metals are very rapid, interstitial diffusers in GaAs, although their final situation is substitutional, and this has led to their almost total disuse from GaAs technology.

As an example, diffusion of Cr has been investigated by several groups, (unfortunately, with disparate results). Diffusion into GaAs from a surface source (indiffusion) produces diffusion rates that are larger than those from within the host lattice. However, there are clear surface effects that result from a chemical reaction between Cr and GaAs. After diffusion the profile is deep, indicating interstitial diffusion, but nearly all the Cr atoms seem to act as acceptors, i. e. they are substitutional [23.54]. Thus, there are similarities with the earlier case of Zn diffusion. This interstitial–substitutional diffusion is probably common in GaAs, although details are rare. One case where some justification for this mechanism was obtained was the case of Fe diffusion. Here, electron paramagnetic resonance (EPR) studies of the diffused layer showed good agreement between $\text{Fe}_{\text{Ga}}^{3+}$ and the atomic concentration of Fe determined by radio-tracer measurements although the rapid penetration rate strongly indicated interstitial diffusion [23.60].

23.4 Ion Implantation into GaAs

Modification to the electronic properties of the near-surface region of a semiconductor is the basis of nearly all devices. In Si this is routinely performed by diffusion or ion implantation. However, most microwave

GaAs devices that require this modification are n-type devices such as metal–semiconductor field-effect transistors (MESFET) and, as mentioned previously, diffusion of shallow donors into GaAs is not technically

possible. MESFETs are majority carrier, n-type conduction devices that are fabricated on the surface of SI GaAs. They need abrupt depth profiles of two carrier concentrations (Sect. 23.11.2 for more details). In ion-implanted MESFETs, these regions are produced by direct implantation of donor atoms into the surface. The depth profile of implanted ions can be calculated using computer programs. In general, the greater the energy of the ion, the greater is the mean penetration and the greater the mass of the ion, the lower the penetration.

Unfortunately, damage to the host crystalline structure is associated with the implantation, resulting in zero electrical activity from the implanted ions. This damage must be repaired by thermal annealing. Two types of anneal schedule are used, furnace annealing and rapid thermal annealing (RTA). The first entails heating the GaAs in a conventional furnace at 850–900 °C for at least 15 min. RTA uses a higher temperature of 900–950 °C but for times of only approximately ten seconds. In each case the surface must be protected by an impervious layer to stop the loss of As. This is

often silicon nitride (Si_3N_4) but the simple expedient of annealing two wafers face to face will reduce As loss to insignificant amounts. The recovery of electrical activation, being a thermally activated process, has an associated activation energy that is usually less than 1 eV. Because that for diffusion often exceeds 2.5 eV, careful implant annealing does not result in appreciable diffusion. A full list of these parameters is given in the article by Sealy [23.61].

In Si, implant anneals remove nearly all this damage; in GaAs only a fraction of the implant damage can be removed, especially after high implant fluences, unless the heat treatments are applied for times and temperatures that are incompatible with device manufacture. Anneals of acceptor-implanted GaAs are more effective with nearly 100% activation of the implant being achievable. The inefficient activation of donor implants in GaAs is not fully understood.

Applications of ion implantation, mostly reserved for MESFETs and monolithic microwave integrated circuits (MMICs) based on MESFETs, are discussed later.

23.5 Crystalline Defects in GaAs

Crystal defects are crucial in determining many electrical and optical properties of GaAs. Most of these defects are incorporated during growth into melt-grown (sub-

strate) material but they can also be introduced in later processing steps.

23.5.1 Defects in Melt-Grown GaAs

Structural Defects

Previously, the problems associated with melt growth were introduced. The growing crystal, cooling from the outside, experiences compressional stresses that cause plastic deformation by the introduction and subsequent motion of dislocations.

In SI GaAs the polygonised dislocation structure is complex. Such structure is best revealed by specialised chemical etching of the surface to reveal dislocation arrangements [23.62, 63] or reflection X-ray topography [23.64]. Each technique reveals classical dislocation cells, *lineage* (the boundary between sections of the crystal that have a small tilt misorientation and caused by dislocation motion) and slip. The lineage occurs preferentially along the (110) diameters. A typical image from a 20-mm² area from a 3-inch-diameter (001) SI LEC GaAs is shown in Fig. 23.15.

The use of selective etching by molten alkalis is commonly used by substrate manufacturers in order

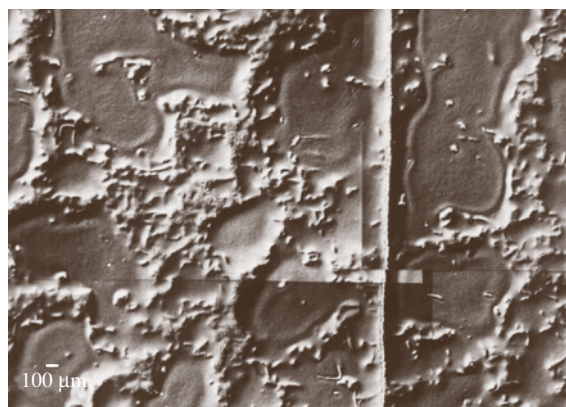


Fig. 23.15 A 20-mm² area of a 3-inch-diameter {001} SI LEC GaAs after etching to reveal dislocations. Many of the dislocations are arranged in a cellular wall formation enclosing regions where the dislocation density is small. The nearly vertical line, closely aligned along [110] is lineage where dislocations have interacted to produce a small-angle tilt boundary. (Courtesy of D. J. Stirland)

to reveal dislocations passing through the surface and being revealed as pits. Thus, they obtain an *etch pit density* (EPD), which corresponds to the dislocation density [23.65]. However, transmission electron microscopy (TEM) is a preferable, though much slower, method for investigating dislocations [23.66].

VGF GaAs contains a much lower density of dislocations but even these are found to be in the form of a cell structure [23.68]. n^+ -GaAs is much less dislocated and no cell structure is present. It is thought that the slip velocity of dislocations on their glide planes is reduced in highly n-type material, so that n^+ -GaAs contains far fewer dislocations than SI or p-type material. However, this cannot be the only reason for this hardening process as the addition of neutral atoms such as In at high concentration can also reduce the dislocation density to near zero [23.69]. The dislocation structures also contain microscopic precipitates, which have been found to be hexagonal, elemental As [23.30, 70]. These are absent in n^+ -GaAs.

Point Defects

SI GaAs is a relatively pure material. It contains boron impurities at high concentration but these are neutral (Sect. 23.1.2). Silicon and sulphur, both shallow donors, are found at concentrations of around 10^{15} cm^{-3} . Concentrations of other electrically active impurities are extremely low and can be ignored.

The exception is carbon, a shallow acceptor, which occurs naturally in LEC GaAs, but which is often added intentionally to VGF GaAs, at concentrations of $10^{14} - 10^{16} \text{ cm}^{-3}$. For this reason carbon must be treated as a dopant and not an impurity. The accurate measurement of the carbon concentration is challenging. The standard method uses the low-temperature far-infrared absorption due to the localised vibrational modes (LVM) of carbon acceptors, a technique that has a sensitivity of around 10^{14} cm^{-3} [23.71, 72].

However, SI GaAs contains many native defects, and these are listed in Table 23.2.

Most native point defects are deep donors, at concentrations of $10^{13} - 10^{14} \text{ cm}^{-3}$ with one, EL2, the As antisite defect (As_{Ga}), being dominant, existing at a concentration of $1 \times 10^{16} - 1.5 \times 10^{16} \text{ cm}^{-3}$. Compensation of the carbon acceptors by EL2 pins the Fermi level near the centre of the band gap, the other point defects being fully ionised by this process. The lack of native acceptors at concentrations above those of donors makes the presence of carbon (or other chemical acceptors) mandatory for SI behaviour.

The EL2 Centre

EL2 is a native deep double donor. Its first ionisation state is at 0.75 eV above the valence band. This is the level that controls the Fermi level in SI GaAs. A second ionisation from + to ++ occurs at an energy at

Table 23.2 Deep electronic levels observed in melt-grown SI GaAs. EL is an electron level and HL is a hole level. In most commercial material only EL2 exists at concentrations exceeding 10^{15} cm^{-3} and is the only deep level assumed to be involved in the compensation mechanism to give SI properties. (From [23.67])

Label	Origin	Concentration (cm^{-3})	Emission energy (eV)	Capture cross section (cm^2)
EL11			$E_c - 0.17$	3×10^{-16}
EL17			$E_c - 0.22$	1.0×10^{-14}
EL14			$E_c - 0.215$	5.2×10^{-16}
EL6	Complex defect	$10^{14} - 10^{16}$	$E_c - 0.35$	1.5×10^{-13}
EL5		$10^{14} - 10^{16}$	$E_c - 0.42$	10^{-13}
EL3		$10^{13} - 10^{15}$	$E_c - 0.575$	1.2×10^{-13}
EL2	Native defect As_{Ga} or $[\text{As}_{\text{Ga}} - \text{X}]$	$5 \times 10^{15} - 3 \times 10^{16}$	$E_c - 0.825$	1.2×10^{-13}
HL10		Below 2×10^{14}	$E_v + 0.83$	1.7×10^{-13}
HL9			$E_v + 0.69$	1.1×10^{-13}
HL7		1.7×10^{15}	$E_v + 0.35$	6.4×10^{-15}
"Ga _{As} "	Gallium antisite Ga_{As} or boron antisite B_{As}	$3 \times 10^{15} - 3 \times 10^{16}$ (dependent on Ga richness of melt)	$E_v + 0.077$ $E_v + 0.203$ (double acceptor)	

0.54 eV above the valence band. Only the singly ionised state is paramagnetic but EPR measurements on this state demonstrated that EL2 contained the As antisite defect [23.73].

EL2 exhibits a broad near-infrared absorption band in SI GaAs, which has been separated into components from neutral and singly ionised EL2 [23.74, 75].

EL2 has come under extreme scrutiny over the past two decades, not only because it results in SI behaviour, but also because it exhibits a remarkable photo-quenching property; at low temperatures EL2 defects can be excited into an inert metastable state by irradiating the GaAs with sub-band-gap light [23.74]. This process is associated with a photoactivated relaxation of As_{Ga} along a $\langle 111 \rangle$ direction into an interstitial site to produce a $[\text{As}_i - \text{V}_{\text{Ga}}]$ complex. Normal deep-donor behaviour is recovered by warming above 140 K [23.76, 77].

The two states of EL2 are shown in Fig. 23.16. Interestingly, a similar deduction about the motion of donors has been made in highly n-type GaAlAs when the Al content exceeds a critical value. The effect, which, results in the loss of normal donor behaviour, is ascribed to the generation of the DX centre. Thus, conduction electrons are lost on irradiating the cooled sample with sub-band-gap light. Like EL2 in GaAs, the recovery from DX-like behaviour is recovered by warming the sample.

High concentrations of EL2 are found in close association with dislocations in as-grown material. Coupled with the presence of As precipitates, this demonstrates that the environments of dislocations are very As-rich.

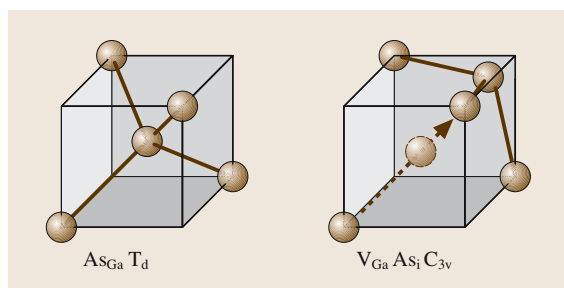


Fig. 23.16 The two states of EL2. The *spheres* represent As atoms. In T_d symmetry the central As atom is tetrahedrally bonded to four nearest neighbour As atoms as As_{Ga} . In the state represented by C_{3v} symmetry this As atom has been excited to move along the $\langle 111 \rangle$ direction where an uncharged metastable site is present. It is now represented by $[\text{V}_{\text{Ga}} - \text{As}_i]$. Heating above 150 K causes the As atom to return to the stable As_{Ga} structure. (After [23.27])

EL2 concentrations are rendered uniform by ingot anneals which, as expected, also improve the uniformity of electrical properties.

Concentrations of EL2 defects are easily determined by measuring their room-temperature infrared absorption at wavelengths near $1 \mu\text{m}$ [23.74]. In practice, EL2 concentrations are always $1.0 \times 10^{16} - 1.5 \times 10^{16} \text{ cm}^{-3}$ in crystals grown from a melt that is slightly As rich. The use of an infrared-sensitive closed-circuit television (CCTV) camera or photographic film can allow mapping of the concentrations to be made [23.78].

Another defect, often referred to as the *reverse contrast* (RC) defect can also be mapped. This must take place at lower sample temperatures but otherwise uses a similar technique [23.79, 80]. RC absorption also reveals dislocations. However, unlike EL2 concentrations, those of RC defects are not homogenised by these heat treatments, indicating that other point defects may also be resistant to standard anneal protocols. RC defects occur at relative concentrations that are the reverse to EL2 defects in unannealed material (they are high where $[\text{EL2}]$ is low, and vice versa) and seem to control minority carrier lifetime in SI GaAs and, thus, their low-temperature luminescence. Their identity as being As vacancies was demonstrated by positron-annihilation methods [23.81] and their nonuniform distribution was explained, at least in part, by the influence of dislocations on the As Frenkel reaction on the As sublattice as the crystal cools [23.82]. The identities of other point defects in SI GaAs have not been determined.

23.5.2 Epitaxial GaAs (not Low Temperature MBE GaAs)

Structural Defects

Because they cannot end within the crystal, dislocations that pierce the surface of the substrate must grow into the epilayer and, for devices that rely on minority carriers, this is a problem. The reason for this is that dislocations in GaAs can act as potent minority-carrier recombination centers. This may result from the presence of a high density of electrically active dangling bonds at the dislocation core. However, enhanced recombination from the presence of a relatively high concentration of point defects around the core (Cotterell atmosphere) cannot be totally ruled out. The latter would seem to be less likely in epitaxial GaAs because of the low growth temperatures and the low measured concentrations of these defects. However, the types of atoms that are expected to act as recombination centers are also expected to be rapid diffusers and their concentrations around dislo-

cations could be much greater than is estimated. The deleterious effects of these *threading dislocations* are the prime reason for the development of VGF and similar substrates.

Point Defects

Epitaxial GaAs contains a far lower concentration of intrinsic point defects than melt-grown material because of the low temperatures to which these materials are subjected. EL2 defects are found in epilayers grown by all the VPE techniques at concentrations up to 10^{14} cm^{-3} making it both the most common trap and the one occurring at highest concentration. Higher concentrations may be generated after growth by annealing at temperatures near 900°C . In this way, lightly p-type VPE GaAs can be rendered SI, see Sect. 23.2.2. A comprehensive list of electron traps has been presented in [23.83] and a similar list of hole traps has been presented in [23.84]. Capture cross-section determinations have been presented in [23.85].

In general, most point defects in epitaxial material are extrinsic and result from impurities present in the precursors or in the growth apparatus. In modern equipment, using purified starting chemicals, residual impurities are typically shallow donors such as Si and S that are difficult to remove. These produce a residual carrier concentration of around 10^{14} cm^{-3} in the best commercial, nominally undoped, epitaxial GaAs. There are notable exceptions to the high purity usually found in MOCVD and MOMBE GaAs

when carbon is used as a p-type dopant. Hydrogen can enter the growing layer and passivate the acceptors, see [23.86, 87], a case where intentional doping by one species can result in unintentional incorporation of an impurity, see also the previous case of Si in LEC GaAs. The hydrogen concentration is up to 50% of the carbon concentration in MOMBE GaAs and up to 20% in MOCVD GaAs. Fortunately, it can be removed by a post-growth heat treatment at 600°C for around 10 min.

23.5.3 LTMBE GaAs

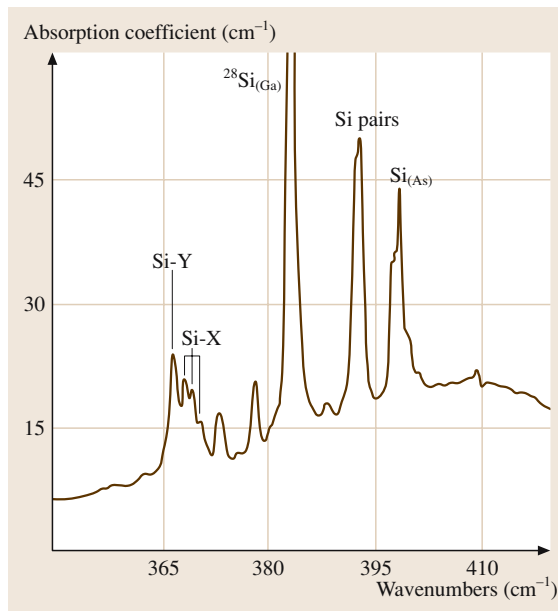
The low growth temperature and an excess of As cause this material to be very As-rich [23.47, 88]. Initially the excess As is in solution and device processing causes it to precipitate. For reasons of reproducibility, an anneal at around 500°C is applied before processing which stabilises the LTMBE GaAs, which now contains a very high number of As precipitates and increases its resistivity to over $10^8 \Omega \text{ cm}$ [23.89]. The high resistivity is probably due to high concentrations of point defects, especially EL2, but Schottky barrier effects, resulting from the As/GaAs interfaces, which would result in virtually all the material being depleted, cannot be ruled out [23.90]. The acceptor species, which must be present to ionise the EL2 defects and pin the Fermi level, is widely believed to be the Ga vacancy, V_{Ga} [23.91] and there is some positron-annihilation spectroscopy evidence to support this [23.92].

23.6 Impurity and Defect Analysis of GaAs (Chemical)

Concentrations of impurities in all grades of GaAs, with the exceptions mentioned previously, are extremely low and always less than 1 part per million (ppm) or $\approx 4 \times 10^{16} \text{ cm}^{-3}$. Secondary-ion mass spectrometry (SIMS), or similar techniques like glow-discharge mass spectrometry (GDMS), are the chosen method for detecting most impurity elements, see [23.93]. In these methods, the surface is ablated and ionised. The secondary ions from the material under investigation are then passed through a mass spectrometer before detection. Sensitivities of better than 10^{14} cm^{-3} are quoted by most SIMS and GDMS laboratories. Unfortunately, important species like hydrogen, oxygen and carbon are not detected with adequate sensitivity by these techniques unless considerable care is taken to remove these impurities from the vacuum ambient of the mass spec-

trometer. This normally requires extra pumping often by cryogenic means (cooling a section of the instrument to near liquid-helium temperatures to freeze impurities and remove them from the vacuum), atypical extra components of most spectrometers. Because impurities and their concentrations tend to be specific to manufacturers and GaAs is being improved continuously, an example of a chemical analysis is not given.

At low concentrations, C, H, and O are difficult to measure in epitaxial GaAs because of the low volume of material that is available for analysis. In bulk GaAs, C, and H are best detected using infrared absorption by localised vibrational modes (LVM). This technique exploits the fact that light impurities in a solid have vibrational frequencies that do not couple well to the lattice modes (phonons); they vibrate at frequencies above



the maximum optical-phonon frequency (Reststrahlen) and lose this energy only slowly to the lattice. This gives them a long lifetime and a narrow bandwidth, resulting in sharp absorption bands which fall in the mid- to far-infrared part of the spectrum at 4–30 μm . Because vibrational frequencies and their absorption strengths depend on the atomic weights of the atoms and their isotopic abundances, respectively, LVM measurements can reveal the atom type unambiguously in many circumstances [23.72].

Measurement of these narrow absorption bands, often employing a Fourier-transform infrared (FTIR) spectrophotometer and low sample temperatures, pro-

Fig. 23.17 An LVM absorption spectrum of GaAs doped with Si at a concentration of around $5 \times 10^{18} \text{ cm}^{-3}$. Lines due to Si_{Ga} donors are present at 384, 379 and 373 cm^{-1} , corresponding to vibrations of ^{28}Si , ^{29}Si and ^{30}Si on Ga lattice sites, respectively. Lines from Si_{As} shallow acceptors, [$^{28}\text{Si}_{\text{Ga}} - ^{28}\text{Si}_{\text{As}}$] nearest-neighbour pairs and complexes labelled Si-X and Si-Y, seen only at high Si concentration, are also seen. (After [23.72])

vides sensitive information not only of concentrations but also of the atomic environments of the impurity atoms.

Unfortunately, absorption from free carriers is very strong in this spectral region and this must be reduced, usually by irradiation of the sample with high-energy electrons. At suitably low fluences, this produces deep levels that compensate the material, apparently without otherwise affecting the impurities. However, this makes the technique difficult for routine measurements. SI GaAs does not need this treatment making it the method of choice for the measurement of carbon concentration (Sect. 23.5).

A LVM absorption spectrum of GaAs containing a high concentration of Si is shown in Fig. 23.17.

The importance of oxygen in commercial GaAs has been questioned for many years. However, the maximum concentrations detected in bulk material are usually below the sensitivity limit for chemical techniques or LVM absorption. The only reports of sizeable concentrations of O in LEC GaAs were based on few, carefully prepared crystals [23.94].

Other methods of chemical analysis are either in disuse or out of the scope of this short section and will not be discussed here.

23.7 Impurity and Defect Analysis of GaAs (Electrical)

23.7.1 Introduction to Electrical Analysis of Defects in GaAs

One of the simplest ways of assessing a sample of GaAs is a measurement of carrier concentration and mobility. This is easily achieved in conducting material by Hall effect analysis. Applications of the Hall effect to GaAs have been discussed in detail by Look in [23.95]. This rather simple measurement can give considerable information about concentrations of shallow and deeper electronic levels. However, its sensitivity to the latter is rather limited. Also measurements in SI GaAs are

complicated by the difficulty in passing sufficient current through the material to achieve a measurable Hall voltage and the extremely high source impedance of the latter. Fortunately, there exist other highly sensitive electrical assessment techniques that can be used.

However, before these are introduced the importance of mapping needs to be emphasised. It is clearly important to reveal changes of electrical parameters across a wafer and changes that may occur from wafer to wafer. These measurements must be performed rapidly and preferably without contact to the GaAs wafer. Such measurements are often limited to resistivity and as-

sume that this parameter does not vary with depth. For conducting materials, microwave eddy-current loss measurements can be made very rapidly with a resolution of a few mm. These mappers have been available commercially for many years. For SI GaAs the absorption of microwave radiation is very small and another technique, time-domain charge measurement (TDCM), has been developed [23.96]. This method, which measures the dielectric relaxation time in order to give resistivity data, can have a spatial resolution of a few tens of μm [23.97].

Shallow-Level Defects

In p-type GaAs shallow acceptors have an appreciable ionisation energy and in nondegenerate material the resulting free holes can often be frozen out by cooling to temperatures close to 4.2 K. The Hall effect can be very useful here. An Arrhenius plot extracted from carrier concentration measurements taken as a function of temperature can reveal this activation energy and, therefore, the acceptor. The presence of more than one acceptor makes the analysis more difficult; if compensating donors are present they are not revealed directly but mobility data can indicate their concentration.

This method does not work with n-type material because the ionisation energies of shallow donors are very similar and, although total donors concentrations can be established, identification is impossible. Once again, it is assumed that no acceptors act as compensating defects.

However, a more complex technique, photothermal ionisation spectroscopy (PTIS) can identify donors [23.99]. This is based on the fact that the energy, ΔE , to excite the bound s electron to its first excited 2p state depends on the donor type. In PTIS, the sample, with Ohmic contacts, is cooled to temperatures near 10 K. Its conductivity is measured under illumination with light from a monochromator. When the photon energy is equal to ΔE , the 1s electron is excited to the 2p state. It is now ionised very rapidly by thermal excitation to the conduction band where it can be detected as a photocurrent. PTIS is very sensitive (less than 10^{14} cm^{-3} has been cited) but it fails for concentrations much higher than 10^{16} cm^{-3} because overlap of the electron wavefunctions causes broadening of the 1s–2p transitions. Extra information can be obtained by application of a magnetic field, stress or other perturbations to the crystal and these are discussed in [23.100].

Deep-Level Defects

Deep levels are produced by many impurities and intrinsic point defects. In many ways, these are more

insidious than shallow levels as their presence is often only detected when fabricated devices fail to perform adequately. Deep levels can result in unexpected noise, parameter drifting, electrical hysteresis, poor light output from optoelectronic devices, etc. These effects arise from carrier trapping and subsequent de-trapping at deep levels, and their role as recombination–generation (RG) centres. Detection and measurement depends on whether the GaAs is SI or conducting.

SI GaAs. Deep levels can trap carriers, produced optically by illumination with above-band-gap light, and then release them thermally. The released carriers, electrons or holes, can produce a current in an external circuit. The temperatures at which different traps release their carriers can be used to determine their ionisation energies. This is the basis of thermally stimulated current (TSC) spectroscopy, a technique that is especially useful for assessing high-resistivity materials such as SI GaAs. See [23.101] for discussions of this and other electrical assessment techniques. Two Ohmic contacts are placed on the sample. After optical filling of the traps at low temperature and at zero applied bias, the temperature is increased at a constant rate with the GaAs under bias and a sensitive current meter in series with the voltage supply. The current in the external circuit shows a series of peaks corresponding to trap emptying, the deeper traps requiring higher temperatures before they can con-

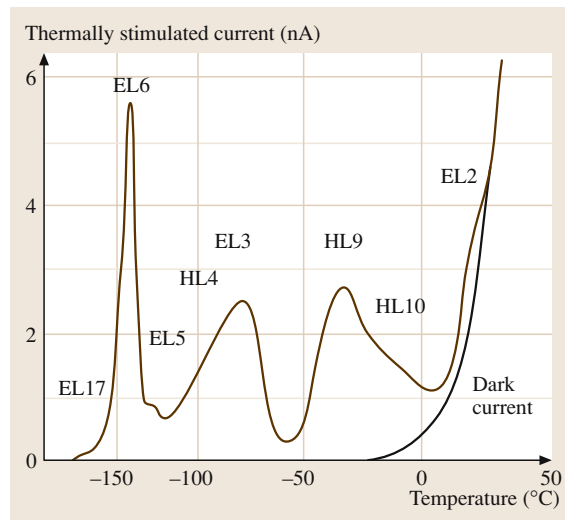


Fig. 23.18 TSC spectra recorded on a 350- μm -thick sample of SI GaAs. The traps have been filled by illumination with white light for 3 min at 80 K. The reverse bias is 25 V. (After [23.98])

tribute. This technique is sensitive (10^{13} cm^{-3}) and can be semiquantitative. It cannot, however, show whether a deep level is an electron or a hole trap, because the emptying of either will produce the same external current. A typical TSC spectrum from a sample of SI GaAs is shown in Fig. 23.18.

Conducting GaAs. Application of TSC to conducting semiconductors with Ohmic contacts is impracticable because of the high temperature-independent background current that would flow. However, the filling and emptying of traps is still a useful approach. If a rectifying junction can be formed, applying a forward bias can fill traps; the subsequent application of a reverse bias produces a depletion layer in which the traps can empty thermally. The emptying of each trap type produces a flow of charge that changes exponentially with time so that each current pulse is a sum of several exponentials with different time constants.

Although the current in an external circuit could be used to perform the subsequent measurement, it is easily shown that more information can be obtained from the change in depletion-layer capacitance because this can reveal whether electrons or holes are being de-trapped. If this capacitance change transient, also a sum of exponentials, is analysed by applying the *rate window* concept [23.103], the data can be presented as a spectrum of capacitance change in that window versus temperature. This is known as deep-level transient spectroscopy (DLTS) and a typical spectrum from a sample of GaAs is shown in Fig. 23.19.

DLTS is quantitative, giving defect concentrations, ionisation energies, capture cross sections and the identification of the type of trap (electron or hole). The labelling of the deep levels in Table 23.2 followed the early application of DLTS to GaAs.

DLTS equipment includes a sensitive and fast capacitance meter, usually furnished by Boonton, Inc., and a control box that applies the forward and reverse bias voltages as well as applying the rate window analyser to the output. The sample must be placed in a variable-temperature cryostat. As a rule, DLTS is controlled by computer. Often the fabrication of a p–n junction is inconvenient and a Schottky contact is used. Because minority carriers cannot be introduced into a semiconductor by this means, DLTS will only be sensitive to majority-carrier traps. If minority traps are to be investigated using the Schottky barrier method, minority

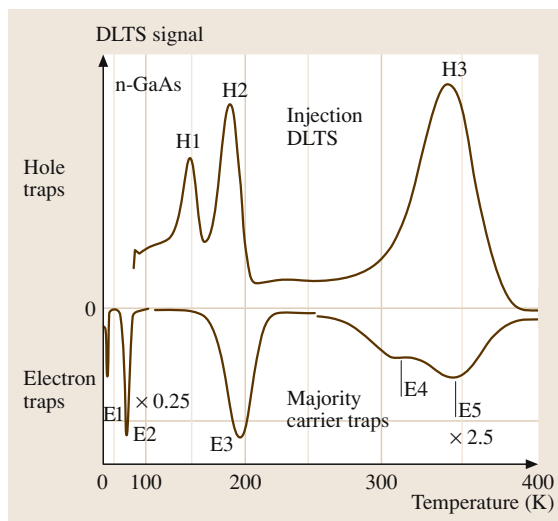


Fig. 23.19 Minority and majority traps introduced by high-energy electron irradiation of a sample of n-type GaAs after the fabrication of a p⁺–n diode. After [23.102]. For further details [23.101]

carriers must be introduced using a pulse of above-band-gap light instead of forward-bias, optical DLTS (ODLTS), Fig. 23.19. The uses of the many variations of DLTS have been discussed and assessed by several workers [23.91, 95, 101].

The sensitivity of DLTS depends on the carrier concentration, which, for standard operation, must exceed the total deep-level concentration by more than an order of magnitude. Within this restriction, the sensitivity is about two orders of magnitude less than the carrier concentration, i.e. for undoped LPE GaAs with a carrier concentration of 10^{16} cm^{-3} , DLTS will be able to measure defects down to the 10^{14} cm^{-3} range.

Although DLTS is widely used, it has a serious disadvantage: each trap produces a wide peak and several traps of similar emission properties cannot be distinguished. So-called Laplace-transform DLTS has been developed to overcome this [23.104]. A rate window is no longer used but the transient is analysed directly in terms of its exponential components by applying an inverse Laplace transform. However, this is demanding in terms of temperature control and the linearity of capacitance measurement. Older techniques such as photo-capacitance spectroscopy are beyond the scope of this section but are reviewed in the books mentioned above.

23.8 Impurity and Defect Analysis of GaAs (Optical)

23.8.1 Optical Analysis of Defects in GaAs

For obvious reasons, analytical techniques that do not require electrical contacts to be made to the sample, or are otherwise nondestructive, are attractive to device manufacturers. Many of these are optical in nature. The use and limitations of *LVM* absorption has been addressed in Sect. 23.6 and that for measuring EL2 concentrations was described in Sect. 23.1.3. Other infrared absorption techniques are of limited interest; we concentrate on light emission, or luminescence, methods.

Photoluminescence (PL)

Luminescence in semiconductors results from the radiative recombination of excess carriers. With the exception of direct recombination via free excitons, luminescence takes place via the mediation of defects [23.105, 106]. The energies of luminescence bands give information of the defects, the intensities are related (albeit in a rather indirect way) to concentrations and the total luminescence intensity is related to concentrations of parallel, nonradiative recombination paths. If the excess carriers are produced optically, often by illumination of the sample with a laser beam, the technique is photoluminescence. Excitation using a beam of high-energy electrons in a scanning electron microscope (*SEM*) results in cathodoluminescence (*CL*). Although the scanning ability of the *SEM* allows *CL* to map the luminescence of the material, *CL* is rarely applied to GaAs.

In either method, excess carriers are produced within the top few μm of the surface. However, this is not the only volume that is probed by the technique, as these carriers will diffuse about one diffusion length, L_D , before they recombine. L_D will be greater in material where the minority carrier lifetime is greater; in general, this will be associated with better, more luminescent material. It follows that there are restrictions on the spatial information that can be obtained, better resolution being obtained with material of lower L_D . Moreover, excess carriers that diffuse to the surface are lost to the measurement as they recombine there without emission of light. This process is quantified by a *surface recombination velocity*, which increases strongly as the temperature is raised. *PL* and *CL* are usually performed at low sample temperature because this reduces thermal broadening of the emission lines and reduces the diffusion of excess carriers to the sample surface. The emitted light is passed through a high-resolution spectrometer before detection by either a suitable photomultiplier tube or semicon-

ductor (usually Si or Ge) detector. A Ge detector must be cooled, although cooling of all detectors generally results in less noise.

Low-Temperature Luminescence from Point Defects

Low-temperature luminescence from shallow acceptors often dominates the spectra of SI GaAs, the energy of the lines being approximated by $E_G - E_A$, where E_A is the energy of the acceptor relative to the top of the valence band. Because E_A depends on the acceptor species, the presence of a particular acceptor can be determined. Similar luminescence from donors is less easy to differentiate as they have similar values of E_D (where E_D is the energy level relative to the bottom of the conduction band). Luminescence from intrinsic deep levels is weak because recombination is mostly nonradiative. However, that from EL2 centres can be resolved both at low and at room temperature [23.98, 107]. The situation from impurity levels, especially transition metals, usually resulting from contamination is different with strong luminescence resulting from many. Of interest is the line due to Cu at 1.05 eV, which is particularly strong and well resolved and allows this common contaminant to be detected. The *PL* spectra from GaAs depend very sensitively on the method of growth, subsequent treatments and the presence of impurities. Readers interested in this technique are directed towards the relevant chapters of [23.1].

Room-Temperature Luminescence Measurements

The broadening effects of temperature and the increased influence of the surface greatly reduces the use of room-temperature luminescence as an analytical tool. However, mapping at room temperature is often attractive because of its relative ease compared to low-temperature studies. It is then a truly nondestructive method giving spatial information of surface properties (that control the surface recombination velocity) and it is often used for assessing substrates and the uniformity and quality of quantum wells (*QW*). A review of these techniques has been given by [23.108].

Mapping of Surface Properties. The surface quality of GaAs substrates is critical to successful epi-growth. Manufacturers perform special cleaning and oxidation procedures to ensure reproducible surface properties, those being suitable for immediate epi-growth being supplied as *epitaxial-ready*. A valuable check on these

properties is scanning PL as changes to the surface result in changes to the surface recombination velocity. Of course, this assumes that the material is well annealed and that residual nonuniformities in point-defect concentrations do not otherwise affect the recombination kinetics, a situation that is never fully achieved. There is no need for a spectrometer in this measurement as the total PL efficiency is altered by the surface and, in any case, PL line widths are thermally broadened. Collection of all the luminescence means that a relatively high signal is detected, with a concomitant increase in the measurement rate. Small-scale structures are often associated with residual volume nonuniformities (see, for example, [23.109]) whilst large-scale effects are most often due to polishing and subsequent contamination of the surface [23.110].

23.9 Assessment of Complex Heterostructures

Most electrically active GaAs, as opposed to substrate material, is used in complex device structures consisting of several layers. In many optical and microwave devices these layers will be of GaAs and $\text{Ga}_{1-x}\text{Al}_x\text{As}$, where x may take several values and the carrier concentration and thickness of each layer is critical.

23.9.1 Carrier Concentration Measurements in Heterostructures

Although Hall-effect measurements accompanied by a process where the surface is gradually removed by chemical dissolution can be used to assess material of more than a single layer, more complex structures normally use capacitance–voltage (CV) profiling. This method is sensitive and has good thickness resolution. However, it does not give carrier mobility information. Some point-defect information can be derived from the frequency dependence of the data, although expert analysis is required. CV measurements are normally a part of DLTS instrumentation and suitable software is then available to analyse the data see [23.93, 101]. In some situations, where individual layer thicknesses exceed a micron or so, the wafer can be cleaved (normally to reveal a {110} surface) and etched in a solution that stains the layers according to doping type and material. Such a cleaved and stained surface can be viewed in an optical microscope and probed electrically for rapid assessment [23.93].

Mapping of Quantum-Well Properties. Quantum wells act as very efficient collectors of excess carriers and are, in general, highly efficient radiative centers. This is the reason for their use in lasers, LEDs, etc. The reproducibility of QW characteristics, especially in terms of QW widths and the quality of the barriers, across large substrates and from wafer to wafer is a key to successful device manufacture and a challenge to all epitaxial growth techniques. Two-dimensional mapping of room-temperature PL using a spectrometer to resolve the spectrum is used to assess the uniformity of these properties. Although the time to map a wafer depends on the spatial and spectral resolution that is required, the efficient radiative recombination at QWs and the bright luminescence makes this a relatively rapid process [23.111].

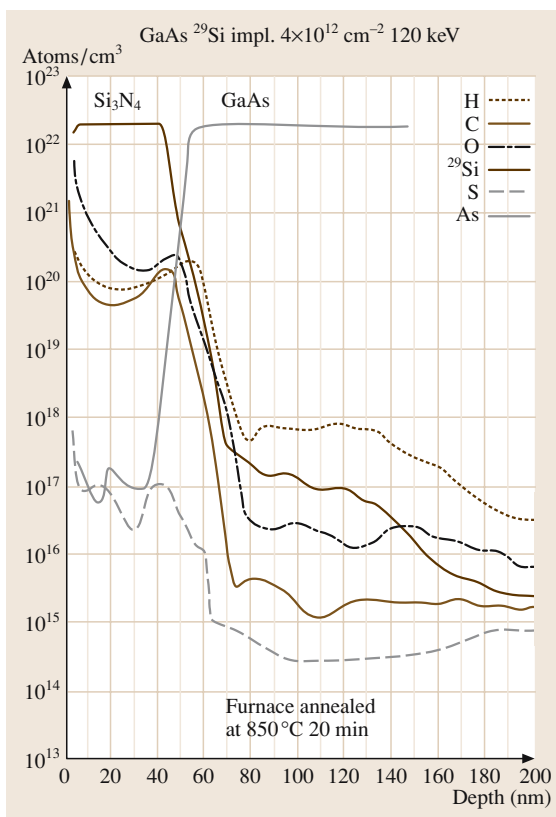


Fig. 23.20 H, O, Si, S, and As depth profiles in reactive-ion-beam sputtered Si_3N_4 film on GaAs as determined by SIMS. (After [23.112])

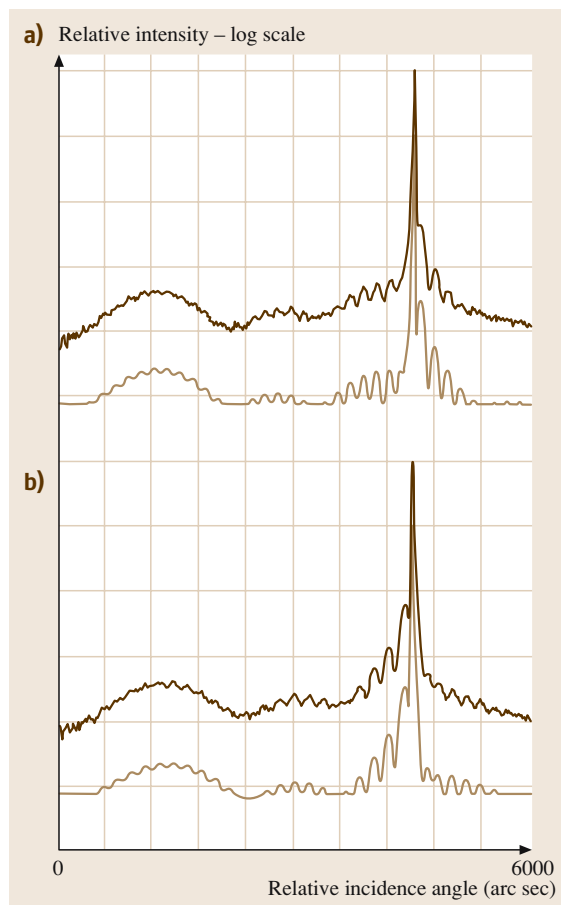
Fig. 23.21a,b Experimental and theoretical rocking curves from a 165-Å-thick layer of $\text{In}_{0.18}\text{Ga}_{0.82}\text{As}$ on GaAs, (a) at the wafer centre and (b) near the wafer periphery. The vertical scales are logarithmic. Note the complexity of these curves for this relatively simple system

23.9.2 Layer Thickness and Composition Measurements

Layer thickness and composition are normally measured together. As an example, dynamic SIMS, where the atomic composition and layer thickness is measured as the ablation of the surface takes place, gives this information effectively, at least for the first few microns of depth. At greater depths, the flatness of the floor of the ablated material degrades. Except that SIMS is destructive, it is often the method of choice [23.93]. Spatial mapping is possible in principle but the increase in machine time makes this prohibitively expensive as a routine procedure. A SIMS profile of a Si_3N_4 film on GaAs, demonstrating the depth resolution and large sensitivity range, is shown in Fig. 23.20.

A nondestructive approach is that of X-ray diffraction, normally double-crystal X-ray diffraction (DXD). X-rays are emitted from a tube and then conditioned by diffraction from a first crystal of high-structural-quality GaAs. The radiation is then diffracted by the sample, which can be tilted to move it in and out of the diffraction condition. A scintillation counter or semiconductor detector detects the diffracted X-rays.

Only the symmetrically diffracted beam is normally used for routine assessment, although other diffraction spots can be more sensitive to the surface structure as the penetration of the beam can be less in these cases. The diffracted energy as a function of tilt is presented as a *rocking curve*, which reveals the small dispersion of the total diffracted energy resulting from lattice strain and other effects. Penetration of the X-rays into the top few microns of a complex structure results in a more complicated rocking curve. However, in virtually all cases, most of this diffracted energy comes from the substrate and diffraction from the epilayers results in small additions to the wings of the rocking curve. It follows that the curve has a large dynamic range of several orders of magnitude and that it can take many hours to obtain the entire curve. Analysis of the rocking curve requires



considerable computation so that the entire process is time-consuming. Often the rocking curve is compared to a computer simulation so that the calculation has to be performed only once. For reviews of these techniques, see [23.113, 114].

Spatial mapping of the rocking curves from substrates to map changes in lattice constant and residual strain is possible, as the rocking curves are simple. However, mapping from complex structures is extremely time-consuming. Figure 23.21 shows rocking curves from a 165-Å-thick layer of $\text{In}_{0.18}\text{Ga}_{0.82}\text{As}$ on GaAs, (a) at the wafer centre and (b) near the wafer periphery. The lower curve of each set of data is a theoretical fit to the experimental curve above.

23.10 Electrical Contacts to GaAs

It is clearly necessary to make electrical contacts to a GaAs device. These are of two types, those with a small internal barrier voltage and which do not affect the passage of a current, *Ohmic* contacts, and those whose internal electrical fields result in a sizeable potential difference and which modify the current flow, non-Ohmic or Schottky contacts.

23.10.1 Ohmic Contacts

A true Ohmic contact would exploit the constancy of the Fermi energies in the semiconductor and the metal. Such a situation rarely occurs and the true nature of practical Ohmic contacts often remains unclear. As an example, the most commonly used Ohmic contact to n-type GaAs is the eutectic alloy of Au–12% Ge. This is evaporated onto the surface and then heated to about 400 °C to produce the required electrical properties. It is believed that the heat treatment results in the in-diffusion of Ge to produce an extremely thin n^+ region, although this may not be the only mechanism; electron microscopy shows that complex intermetallic reactions between the Au–As and Ge take place. If Ni is added to the Au–Ge alloy to produce a ternary, a greatly improved contact results, certainly as a result of the Ni reacting with the GaAs surface to produce intermetallic phases with As and possibly by removing the oxide layer. Other contact materials are sometimes used and many of these have been reviewed in [23.95].

Ohmic contacts to p-type GaAs often use evaporated Au–Zn alloys followed by a heat treatment similar to that used for n-type contacts. The in-diffusion of Zn atoms clearly takes place but the removal of oxides on the GaAs surface by the Zn is also important.

23.10.2 Schottky Contacts

The barrier height between most metals and n-type GaAs is found to be nearly constant at 0.6–0.7 V. This high and constant value is a result of the Fermi energy at the surface being controlled by surface electronic states rather than the properties of the metal. Titanium is the most important Schottky contact. It is usually introduced via a TiPtAu alloy. This exploits the high sticking coefficient of Ti to GaAs whilst the Pt acts as a diffusion barrier to the low-electrical-resistance Au metallisation that would otherwise enter the GaAs. TiAl is sometimes a preferred alternative.

Interestingly, Al also produces a Schottky barrier on p-type GaAs with a barrier height of about 0.6 V. Gold was used as a Schottky contact on n-type GaAs with a barrier height of about 0.9 V but, as discussed above, this is degraded by in-diffusion at high temperatures. Contacts used specifically for high-temperature use include some silicides, with WSi_x being commonly used in some microwave devices. Metals for Schottky contacts have been reviewed in [23.115].

23.11 Devices Based on GaAs (Microwave)

23.11.1 The Gunn Diode

The occurrence of secondary conduction-band minima in the E – k diagram of GaAs leads to the possibility of high-energy electrons being excited from the Γ point where they have low effective mass to the X points where their mass is much greater. This is the transferred-electron effect. Although on applying a greater electric field their kinetic energy is increased, their increased mass leads to a fall in velocity. As a result, the current falls. This appears as a negative differential resistance (NDR) region in the current–voltage characteristics [23.116–118]. If an applied bias keeps the current flow in the NDR region, it can be shown that the electron flow will break up into pulses and the cur-

rent in the external circuit will also be pulsed. Under certain circumstances, this results in the production of power at a particular microwave frequency determined mainly by the thickness of the GaAs [23.119]. The device optimised for microwave power production is called the *Gunn diode*, after J. B. Gunn who first observed this effect experimentally. This device is shown in Fig. 23.22.

The device is very simple, consisting solely of a few microns thickness of high-purity n-type ($\approx 10^{15} \text{ cm}^{-3}$) GaAs with more highly doped contacting regions to improve connection to the outside world [23.120]. The Gunn diode can produce microwave emission at frequencies up to 10 GHz and with powers of a few Watts. These devices have uses as portable microwave sources in radar speed detectors, for example.

23.11.2 The Metal–Semiconductor Field-Effect Transistor (MESFET)

As discussed in the introduction, the increased mobility of electrons in GaAs over those in Si seem to offer the prospect of faster devices, especially improved microwave transistors. This turns out to be based on an oversimplistic interpretation as to how these devices really work. However, consider the first, three-terminal GaAs microwave device, the metal–semiconductor field-effect transistor (MESFET) whose structure is presented in Fig. 23.23.

The MESFET is similar to the Junction Field Effect (JFET) in Si technology with the p–n junction gate being replaced by a simple Schottky gate. Ti-based alloys are normally used for the Schottky contact as these adhere well to the surface. Application of a voltage to the gate modulates the width of the conducting channel and hence its conductivity. The channel is connected to two

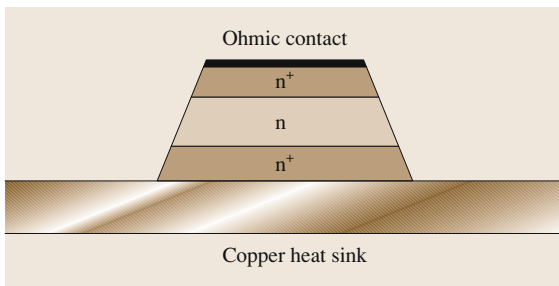


Fig. 23.22 The Gunn diode. The active region is the low-doped n-type GaAs in the centre. (After [23.120])

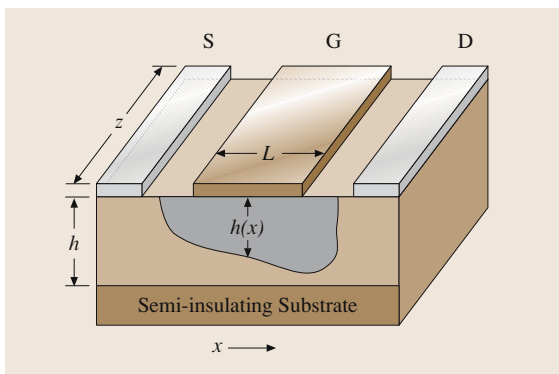


Fig. 23.23 The MESFET. Current flows only in the undepleted region. L is the gate length and determines the speed of the device; z is the gate width and controls the maximum current capabilities of the device; h is the channel thickness. (After [23.121])

highly doped contact regions where ohmic contacts are made. These are the source and drain.

Initially all MESFETs were made by donor ion implantation into SI GaAs substrates using the methods described in Sect. 23.4. However, more recent methods including epitaxial growth by MOCVD and MBE result in higher effective doping levels, more abrupt interfaces, especially at the rear of the channel where a rapid change in carrier concentration is critical and the possibility of depositing alloys such as GaAlAs. Ion implantation is still used throughout the industry, however, especially for digital circuits.

In operation, the gate bias causes the channel to be pinched off. The drain–source voltage produces an extremely high field near the pinched-off channel and the channel current is saturated. Figure 23.24 shows that under these circumstances the electron velocity is expected to be saturated at $\approx 5 \times 10^6$ cm/s, a result of velocity saturation and the transferred-electron effect, the latter producing the reduction in velocity at higher fields.

This value is actually *less* than the saturated electron velocity in Si at similar electric fields and it is clear that the advantages of low electron effective mass at low fields in GaAs are not important. However, microwave measurements show an unexpected extension of high-frequency amplification especially with MESFETs with a small gate length. This is caused by the extremely short

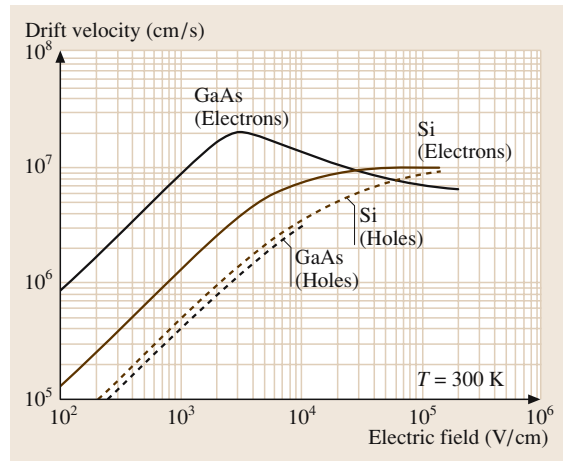


Fig. 23.24 Drift velocity of electrons and holes in Si and GaAs. At low electric fields the carrier velocities increase linearly with electric field (Ohm's law). However, at high fields saturation takes place. The reduction in velocity of electrons in n-type GaAs is a result of the transferred-electron effect. (After [23.4])

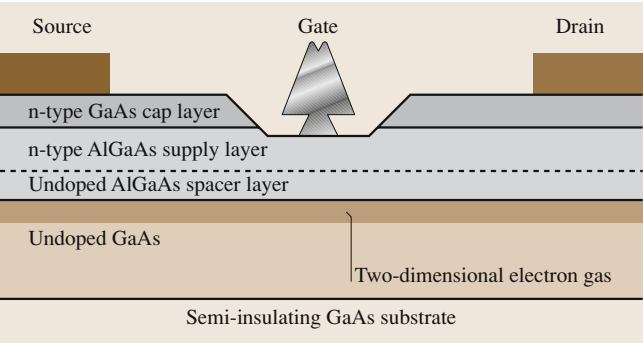


Fig. 23.25 The GaAs/AlGaAs HEMT. (After [23.125])

time that electrons are in the channel, which does not allow them to lose their kinetic energy to the lattice. This effect, known as *velocity overshoot*, allows electrons to reach a velocity near 2.5×10^7 cm/s in short-channel devices and is the first true reason for the high-frequency capability of the GaAs MESFET; the other is the use of a SI substrate. The latter reduces the parasitic capacitances that exist between the device and its contact pads and ground. The cut-off frequency of GaAs MESFETs is inversely proportional to the gate length and is 110 GHz for a gate length of 0.15 μm . The MESFET is discussed in detail in [23.122–124].

23.11.3 The High-Electron-Mobility Transistor (HEMT) or Modulation Doped FET (MODFET)

The channel of a GaAs MESFET must be doped with donor atoms to produce the necessary free carriers. Scattering by the positively charged donor ions results in a reduction in electron mobility and increased noise. In the HEMT the free carriers in the channel are produced by diffusion of electrons from an n-type layer of greater band-gap energy such as GaAlAs.

GaAs cap	40 nm	$4 \times 10^{18} \text{ cm}^{-3}$
Al _{0.23} Ga _{0.77} As	25 nm	$2 \times 10^{18} \text{ cm}^{-3}$
Al _{0.23} Ga _{0.77} As	4 nm	Undoped
In _{0.2} Ga _{0.8} As channel	12 nm	Undoped
GaAs buffer layer	0.5 μm	Undoped
Semi-insulating GaAs substrate		

Fig. 23.26 The layer structure of the AlGaAs/InGaAs pHEMT. (After [23.125])

A diagram of a GaAs/GaAlAs HEMT is shown in Fig. 23.25.

The GaAlAs supply layer of 50 nm thickness, doped n-type to a concentration of about $2 \times 10^{18} \text{ cm}^{-3}$, supplies the free electrons. Alternatively, this layer can be delta-doped with a suitable donor. The undoped GaAlAs layer of about 2 nm thickness reduces Coulombic effects between the donors in the GaAlAs and the electrons in the channel which would otherwise reduce their mobility. The thickness of the undoped GaAs channel layer is not critical and is typically less than 1 μm . Advantages of this HEMT structure are best revealed at low temperature where effects due to phonon scattering are also minimised. However, even at 300 K this device is faster and less noisy than a comparable MESFET. This is because the real improvement in performance comes from the heterojunctions which act to constrain the carriers close to the gate with a resultant increase in high-frequency performance.

One disadvantage of the simple HEMT is the poor transfer of electrons from the GaAlAs to the GaAs channel resulting from the rather small conduction-band discontinuity at the interface. This results in a relatively low electron concentration in the channel and a large residual electron concentration in the supply layer, which acts as a slow GaAlAs MESFET in parallel with the HEMT. For this and other reasons, the simple GaAs/AlGaAs HEMT has been replaced by the pseudomorphic HEMT (pHEMT).

The pHEMT exploits the greater conduction-band discontinuity that exists between GaAlAs and InGaAs. Although there is a large lattice mismatch between GaAlAs and InGaAs, the latter can be grown as a thin, (12-nm) pseudomorphic layer on GaAs. The resulting structure is shown in Fig. 23.26.

Cut-off frequencies up to 300 GHz have been reported for these devices, which can also be designed for use as power devices. There is a strong commercial market for MMICs employing pHEMTs in many microwave applications. These include mobile phones, domestic satellite TV antennas and radar and space use. Again, [23.122–124] offer an excellent description of these device types.

23.11.4 The Heterojunction Bipolar Transistor (HBT)

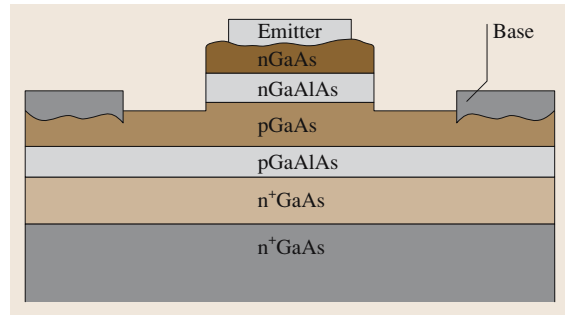
The current gain of a bipolar homojunction transistor is optimised by having an emitter doping that is much higher than the base doping and a base width

Fig. 23.27 Schematic diagram of a heterojunction GaAs/AlGaAs bipolar transistor. The collector contact is made to the rear of the device. (After [23.121])

that is as small as possible [23.121]. Unfortunately, high-frequency operation requires a high doping level in the base to reduce the base resistance. This must be accompanied, therefore, by very heavy emitter doping. However, such heavy doping leads to band-gap shrinkage, resulting in high base injection current and a reduction of the current gain. This argument is true both for npn and pnp homojunction bipolar transistors.

An elegant solution to this problem uses different materials in the device; for an npn device, a heavily doped, n-type, wide-band-gap emitter, a heavily doped p-type base and a lightly doped, n-type wide-band-gap collector. This is the **HBT**.

The base–emitter heterojunction offers a different barrier to hole injection from the base and electron injection from the emitter, resulting in efficient minority-carrier injection into the base [23.126, 127]. The base–collector junction need not be of different materials but a heterojunction may increase the breakdown voltage at the collector. The GaAs/GaAlAs system can be used for these devices, as shown in Fig. 23.27, although **HBTs** based on the alternative GaAs/GaInP



system are superior and have a large market in applications like mobile phones.

Typically, **HBTs** have f_t values of around 40 GHz, although devices with $f_t > 160$ GHz have been reported. They can produce more power per unit area of device than a **MESFET** or **MODFET**. However, catastrophic breakdown can occur at the collector–base junction. High-temperature operation can result in rapid ageing effects, resulting from diffusion of acceptors from the base into the acceptor region. Because of its low diffusion coefficient the use of carbon as the acceptor in the base has reduced these problems. Readers are referred to [23.122, 124] for further reading. **HBTs** find extensive use in mobile telephony, where they are efficient microwave power sources.

23.12 Devices based on GaAs (Electro-optical)

The range of GaAs based electro-optical devices can be separated into emitters, modulators and detectors, the former being of much greater commercial interest than the others.

23.12.1 GaAs Emitters

Light-Emitting Diodes

The direct band gap of GaAs has made it popular for the fabrication of **LEDs** for several decades. The simplest **LED** is a p–n junction in GaAs, which in the early days, was fabricated by diffusing Zn into n-type material. The internal efficiency of this early device was poor at only a few percent but improvements in both material and processing showed that it could be surprisingly high, near 100%, even in such a simple structure. However, the external efficiency is always poor, with most of the emitted radiation being reflected back into the GaAs where eventually it is reabsorbed. This is because of the high

refractive index of GaAs ($n \approx 3.5$), which, for a planar p–n junction, gives a maximum external efficiency from a single face of 2%. There are several approaches to extracting more light from the **LED** and these have been discussed by *Schubert* [23.128]. The use of lenses, antireflection coatings and epoxy domes are only three of the many commercial practices that improve the external efficiency of a simple GaAs p–n junction.

GaAs can be used as a basis for **LEDs** that cover the wavelength range from green (560 nm) through to **IR** wavelengths of over 1000 nm. Table 23.3 is a list of commercial **LEDs** based on GaAs.

The first entry $[(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}]$ is a series of quaternary alloys, all of which have direct band gaps, and lattice-match to GaAs for any value of x . This gives a wide range of colours in the visible part of the optical spectrum. These alloys are grown by **MOCVD** and, although much more expensive than **LEDs** based on simpler materials, have the advantage of being very

Table 23.3 Light-emitting-diode materials and their properties. (After [23.125])

Active region	Gap	Substrate	Lattice match	Wavelength (nm)	Color
$(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$	D	GaAs	Yes	560–640	Green, yellow, orange, red
$[\text{Al}_x\text{Ga}_{1-x}\text{As}-\text{GaAs}]$ MQW ($x < 0.45$)	D	GaAs	Yes	630–870	Red to IR
$\text{Al}_x\text{Ga}_{1-x}\text{As}$ ($x < 0.45$)	D	GaAs	Yes	630–870	Red to IR
$\text{GaAs}_{0.6}\text{P}_{0.4}$	D	GaAs	No	650	Red
GaAs	D	GaAs	Yes	870	IR
GaAs: Zn	D	GaAs	Yes	870–940	IR
GaAs doping SL	I-RS	GaAs	Yes	870–1000	IR
$\text{Ga}_{1-x}\text{In}_x\text{As}$	D	GaAs	Yes	> 870	IR

MQW–Multiple quantum well; SL–superlattice; D–direct gap in k space; I-RS–indirect in real space; IR–infrared

bright and efficient. Light emission from the rear of the device is not possible because the GaAs substrate absorbs it. If a thick layer of GaP is grown as the top layer, a variation of this device can be fabricated where the entire GaAs substrate is removed by etching and the GaP becomes a new substrate. Because the

$(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ is now the material with the lowest band-gap energy, all the light is available for emission from the device and the external efficiency is markedly improved.

The second entry represents multiple quantum wells (MQW) based on GaAs wells with GaAlAs walls and whose confined electron states depend on the well widths. The recombination of excess carriers in the wells efficiently produces light emission, the wavelength of which is engineered by modifying the well widths. These LEDs can emit red light efficiently.

The third entry is for bulk $\text{Al}_x\text{Ga}_{1-x}\text{As}$ where the emitted light wavelength depends on the value of x . In both this and the previous case, x must be less than 45% because at higher fractions of Al the band gap becomes indirect and the efficiency for optical emission falls rapidly. Both these structures allow emission from 630–870 nm to be achieved.

The fourth entry has been mentioned in relation to grading in epitaxial growth. $\text{GaAs}_{1-x}\text{P}_x$ alloys have been the basis for low-cost visible LEDs for many years, those suitable for growing on n-type GaAs, resulting in red emission. Green- and yellow-emitting LEDs are based on GaP substrates. Grading the alloy composition from n-type GaAs until the required n-type $\text{GaAs}_{1-x}\text{P}_x$ alloy is grown produces the correct composition for emission. The next growth is of p-type material of the same composition to produce the p–n junction. The p-type contact is made to the top of this structure to complete the LED.

The next entry relates to the simple GaAs LED while the next demonstrates the use of heavy Zn doping which leads to band-gap narrowing and the emission of light

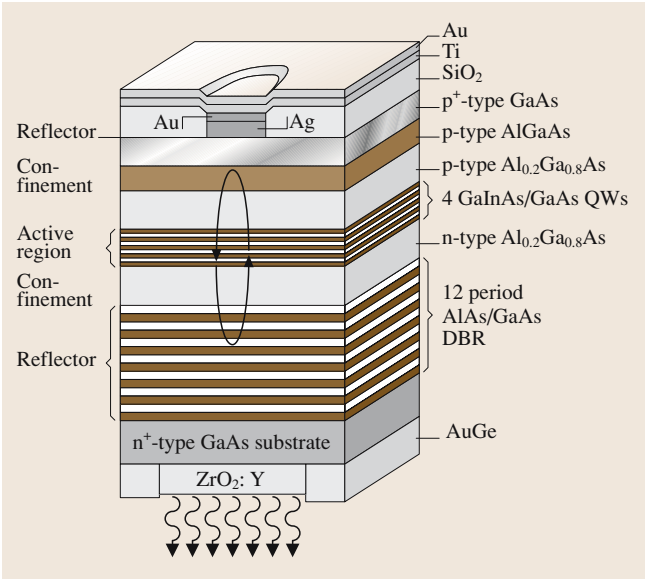


Fig. 23.28 The structure of a GaAs-based RCLED. The active region is a multiple-quantum-well structure. The lower reflector is a distributed Bragg reflector formed by growing a series of AlAs and GaAs layers so that each has an optical thickness of $\lambda/4$ giving near 100% reflectivity at the chosen wavelength. The upper reflector is the simple SiO_2 interface. (After [23.129])

of wavelengths longer than 870 nm. A similar effect has been found by using doping superlattices.

The final entry is an attempt to lower the band-gap energy of GaAs by the addition of In to make $\text{Ga}_{1-x}\text{In}_x\text{As}$ alloys. These alloys are not lattice-matched to GaAs and the introduction of misfit dislocations at the growth interface results in nonradiative processes and a reduction in the efficiency.

Simple IR GaAs LEDs are mass-produced for use in infrared remote controls, optoisolators, emitters for low-cost fibre-optic networks, etc. High-efficiency visible LEDs have a very large market in portable and permanent displays. The former uses include mobile telephones and laptop computers. The second includes advertising and other public displays.

Highly efficient LED structures include the resonant-cavity LED (RCLED) where the emission takes place in a cavity tuned to the emission wavelength. The spontaneous emission along the axis of the resonant cavity is strongly enhanced and the increase in light output in this direction can be more than 10 times that of a more conventional LED. The structure of a RCLED is shown in Fig. 23.28. It should be compared to that of a vertical-cavity surface-emitting laser in the next section.

The recent work dedicated to light-emitting diodes by Schubert is recommended for further reading [23.128].

Lasers

The major difference between an LED and a laser is the existence of a resonant cavity and the confinement of the light in the cavity, which encourages stimulated rather than spontaneous emission in the laser. The use of GaAlAs is very important for the optical confinement as its lower refractive index produces total internal reflection along the axis of the laser. The optical cavity is often produced by simply cleaving parallel {110} facets on the ends of the cavity, the laser being fabricated on a {001} substrate. The high current density that is required to ensure population inversion in the laser is accomplished by careful engineering of the contact metallisation.

There are several types of GaAs laser structure but these can be separated into those where the lasing takes place in bulk material and those where this region is modified by using quantum wells.

Laser-diode structures have been reviewed in [23.130–132] and in many other general textbooks on semiconductors.

The GaAs/AlGaAs Heterojunction Laser. This type of laser is typical of those used in compact disc players,

etc. The structure and operating conditions are shown in Fig. 23.29a and b.

Stimulated emission takes place in the GaAs active region, where excess carriers are injected from the surrounding p- and n-type regions. Different lasers have either GaAs or GaAlAs as these adjacent layers. We dis-

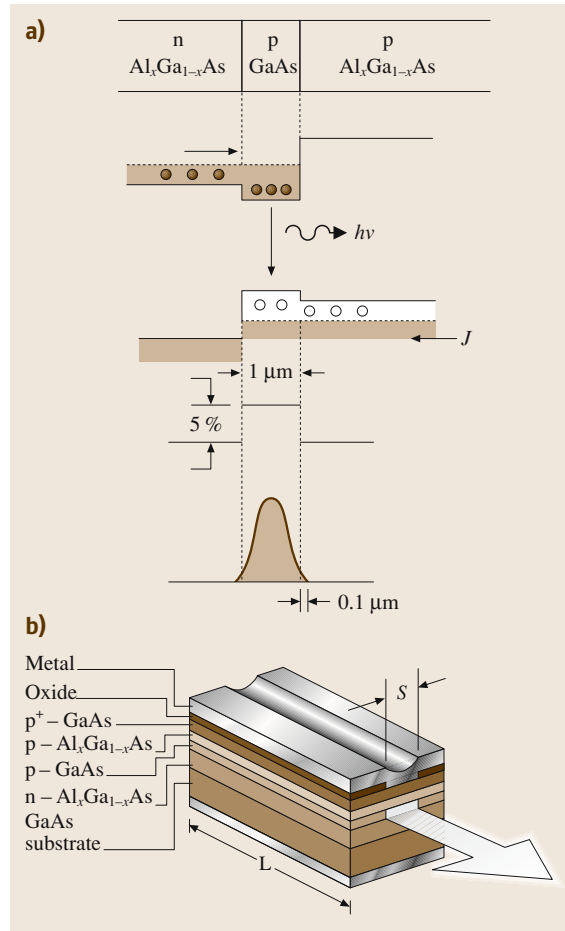


Fig. 23.29 (a) Band-gap diagram of a simple double heterojunction, GaAs/AlGaAs laser. The injected carriers are effectively trapped in the GaAs active region where they recombine to produce light. The increase of refractive index in the active region also constrains the light by total internal reflection. Thus both electrical and optical confinement is achieved. (b) Final structure of a GaAs/AlGaAs ridge laser. The n^+-GaAs substrate and upper p^+-GaAs layers are used for the contacts as metallisation to AlGaAs is less satisfactory. The SiO_2 oxide layer provides further electrical confinement. The optical cavity is defined by the {110} end facets. (After [23.4])

cuss only those lasers with GaAlAs on either side of the active layer, the double heterojunction (DH) laser, as these are most commonly used.

Electrons and holes introduced from the adjacent GaAlAs layers are effectively confined to the GaAs active layer because of the conduction- and valence-band offsets that exist at the heterojunctions. However, this is not optimal for optical confinement and further cladding by $\text{Ga}_{0.4}\text{Al}_{0.6}\text{As}$ is often necessary. This modification results in the device known as the separate confinement heterojunction (SCH) laser. However it is achieved, GaAlAs cladding layers must always surround the active layer to produce optical confinement.

In some designs, the Al content of the GaAlAs cladding layers is graded with the Al content increasing linearly with the distance from the active layer. This graded refractive index (GRIN) structure has optical and electrical advantages and obviates the use of two separate confinement layers. The final structure is often called a GRINSCH laser. A layer of insulating SiO_2 with a thin axial opening to allow electrical connection to the upper p-type layer achieves the electrical confinement. Emission occurs at around 800 nm if GaAs is used in the active layer. If it is replaced by GaAlAs with a low Al content, the emission wavelength can be reduced. Laser structures were originally grown by LPE but large-scale production uses MOCVD or even MBE, although GRINSCH lasers are usually grown by MOCVD. There is a rather high dependence of emission wavelength on temperature with these simple lasers. The use of distributed optical feedback stabilises the output. Though outside the confines of this review, details of these lasers can be found in [23.134].

Although many millions of small GaAs/GaAlAs lasers (less than 20 mW output) are sold annually for domestic applications such as CD players, these lasers can be designed to produce powers exceeding 50 W and are used for pumping Er-doped quartz amplifier fibres for communications use. High-power lasers are not as rugged as smaller units and can suffer from catastrophic failure at the reflecting facets. Special treatments have to be used to prevent these failures.

The Red-Emitting, GaAs/GaAlAs Multi-Quantum-Well (MQW) Laser. The efficient radiative recombination that can occur in quantum wells has been exploited in the design of red lasers based on the GaAs/GaAlAs system. Attempts to produce a red-emitting laser using a conventional active layer fail because GaAlAs with the necessary Al content has an indirect band gap and, therefore, low emission efficiency. However, replacing

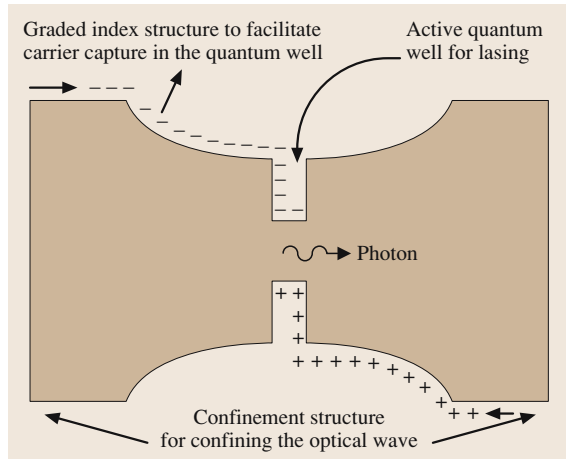


Fig. 23.30 The band structure of a GRINSCH single-QW laser. The large-band-gap material is $\text{Al}_{0.6}\text{Ga}_{0.4}\text{As}$. This is graded to $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ towards the centre of the device, providing electrical and optical confinement. The active region is the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}/\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ quantum well at the centre. The emission wavelength depends on the width of this well, which is typically 1–4 nm. In a GRINSCH MQW laser there is a series of several identical quantum wells. Barriers of typically 5 nm thickness separate them. (After [23.133])

this layer with a series of quantum wells increases the energies of the trapped electrons and holes with respect

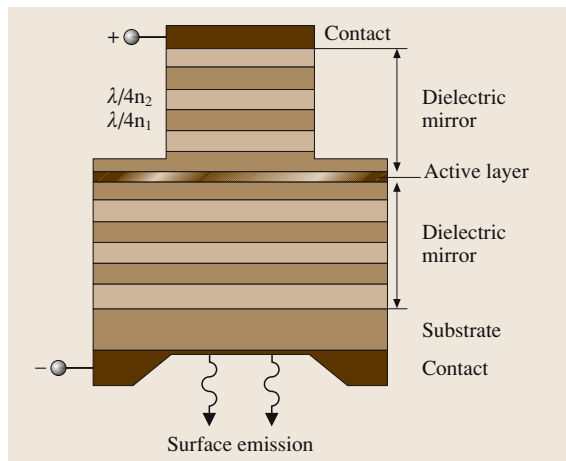


Fig. 23.31 The vertical cavity surface emitting laser (VCSEL). The optical cavity is vertical being defined by the two dielectric mirrors (Distributed Bragg Reflectors). The active region is a multi-quantum well structure at the centre of the device

to the conduction band and the valence band, respectively [23.133]. This effect is the same as used in the **MQW LED**. The structure of a red-emitting **GRIN SCH QW** laser is shown in Fig. 23.30.

These lasers are used in digital versatile disc (**DVD**) players, visible pointing devices, light pens, and several other devices where a visible narrow light beam is important, and where a larger and more expensive He/Ne would have been used previously. The mass production of this complex structure has resulted in the price of these lasers being reduced to the single-dollar range. The **MQW**s in these lasers are very narrow, involving layers that are only a few lattice spacings wide. This is the regime where pseudomorphic growth is possible. Thus, the use of GaAs/GaInAs **MQW**s within GaAlAs cladding layers has produced lasers emitting in the range 0.85–1.0 μm .

The Vertical-Cavity Surface-Emitting Laser (VCSEL). The standard heterojunction laser suffers from several disadvantages from a fibre-optic systems point of view. It is difficult to coat the exit facets with antireflection coatings, it is very difficult to integrate this type of device with microwave amplifiers such as **MESFET**s and the output beam has a cylindrical cross section that is not optimised for injection into a fibre. The **VCSEL** attempts to overcome these problems (Fig. 23.31).

The light is amplified by a **QW** active region as only the **QW** has sufficient gain over such

a thin layer to allow lasing. Even then, the use of simple cleaved mirrors would reduce the overall gain below that necessary for laser operation. Highly reflective multilayer dielectric reflectors (distributed Bragg reflectors often made from a series of GaAs/AlGaAs layers below and above the active layer) are necessary. However, the final device is truly planar, providing a beam of circular cross section. Output powers of several mW have been obtained [23.136] together with integration within a **MESFET** integrated circuit [23.137]. Advanced semiconductor diode lasers and their incorporation within integrated circuits are thoroughly described in [23.137, 138].

23.12.2 GaAs Modulators

It is sometimes necessary to modulate the intensity of a light signal electrically after it has been produced. The near-band-edge absorption of quantum wells can be moved to longer wavelengths by the application of a reverse bias. This is the quantum-confined Stark effect [23.139] and this is the basis of the **MQW** modulator. The modulator structure and the change of near-band-edge absorption with applied voltage are shown in Fig. 23.32.

This device can operate at frequencies up to 40 GHz and has spawned a series of more-advanced devices that are discussed in [23.135].

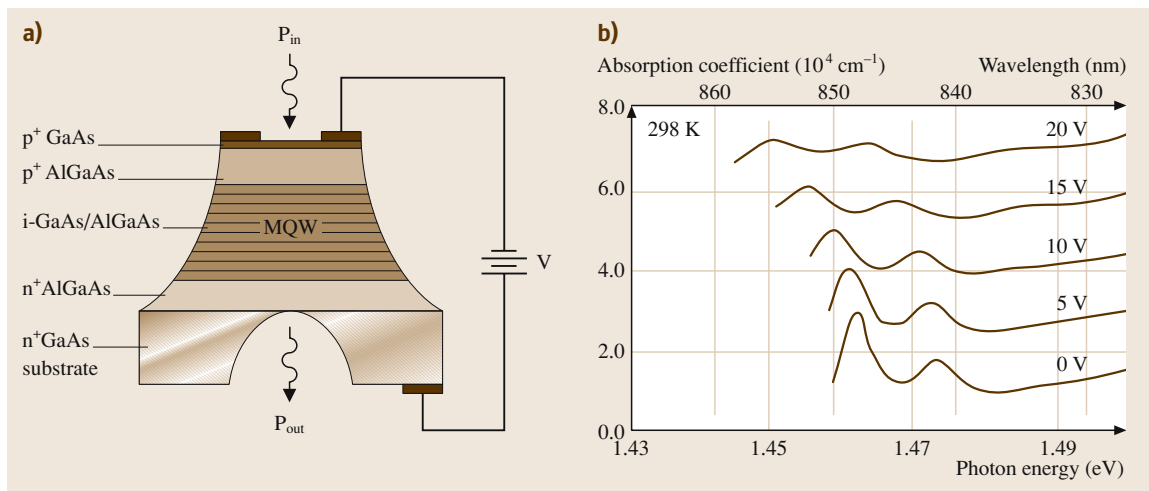


Fig. 23.32a,b The **MQW** optical modulator uses the changes in the near-band-edge absorption spectrum that occurs on the application of an electric field. The device is shown in (a) where it can be noted that the layer structure is very similar to a **MQW LED**. It is, however, operated under reverse bias. (b) Changes in absorption that occur in a **MQW** when a bias is applied. (After [23.135])

23.12.3 GaAs Photodetectors

Solar Cells

As a bulk material GaAs has found great use in solar cells. There are many reasons for this.

1. The band-gap energy is well matched to the solar spectrum and the high optical absorption above the band edge means that most light is absorbed in a thin layer.
2. In extraterrestrial applications, where large numbers of high-energy atomic particles exist, GaAs solar panels damage less easily than Si panels and therefore retain efficiency for longer.
3. In terrestrial use the higher band-gap energy allows GaAs solar cells to operate in concentrator systems (at the focus of a large Fresnel lens which is mechanically aimed at the Sun) where high temperatures are inevitable and Si devices would lose efficiency.

The simplest structure of this device is a simple p-n junction of high-quality GaAs. It is normally covered with a GaAlAs top layer to passivate the surface and to reduce reflection losses. These devices can have an efficiency of over 20% [23.140] Structures that are more

complex involve up to three solar cells grown in series. In a typical two-cell structure, the light sequentially passes through a GaAs solar cell and then a similar GaSb device. Much of the sub-band-gap light that is transmitted by the GaAs is absorbed in the GaSb solar cell ($E_G = 0.72$ eV). Because these solar cells are effectively in series with reducing band-gap energies, more of the sunlight is absorbed than if a single cell were used. The efficiency of these devices can reach 30%. Research and development of multi-junction structures, often using Ge as a substrate and with different layers including GaAs, is very active. Because of their size, especially for space application, solar cells represent a sizeable part of the overall market for GaAs.

Other Photodetectors

The large band-gap energy makes GaAs unsuitable for most optical detector applications; the light generated by GaAs emitters being detected by Si or low-band-gap ternary detectors. However, the ionisation of electrons from bound QW states into continuum states can result in tunable mid- and far-IR detectors. The GaAs/GaAlAs system has been used in this way to make imaging arrays for wavelengths up to $20\mu\text{m}$, see for example [23.141].

23.13 Other Uses for GaAs

There have been attempts to exploit GaAs in other fields. High-energy particle detectors, including those for imaging X-rays, measure the ionisation produced as the particle traverses the detector material. The high stopping power, the large band-gap energy, which reduces leakage current, and the increased radiation hardness of GaAs in comparison with Si would appear to make it suitable for detectors for both scientific and medical use. The ready availability of SI GaAs wafers

has encouraged this work. Unfortunately, these detectors suffer from effects resulting from the presence of deep levels, especially EL2 [23.142]. One application that has shown promise is in high-temperature sensors, amplifiers and processors in automotive, aeronautical and other uses. This exploits the high band-gap energy of GaAs, which allows devices to work at temperatures up to 300°C . An excellent review has been given in [23.143].

23.14 Conclusions

GaAs technology is now mature with devices occupying important niches in the semiconductor marketplace. I have presented the fundamental properties of GaAs and have demonstrated how its unique set of characteristics has allowed it to be exploited in several microwave

and optoelectronic applications. Future developments will probably concentrate on further integration of known devices and improvements to speed, power and reliability rather than the development of novel ones.

References

- 23.1 M. R. Brozel, G. E. Stillman: *Properties of Gallium Arsenide*, 3rd edn. (INSPEC, London 1996)
- 23.2 W. Martienssen: *Landolt-Börnstein*, New Series Group III/ 41Al α , 41A α (Springer, Berlin Heidelberg New York 2001, 2002)
- 23.3 W. Martienssen, H. Warlimont: *Springer Handbook of Condensed Matter and Materials Data* (Springer, Berlin Heidelberg New York 2005) Chap. 4.1, p. 621
- 23.4 S. M. Sze: *Semiconductor Devices, Physics and Technology* (Wiley, Chichester 1985)
- 23.5 W. Koster, B. Thoma: *Z. Met. Kd.* **46**, 291 (1995)
- 23.6 D. T. J. Hurle: A comprehensive thermodynamic analysis of native point defect and dopant solubilities in gallium arsenide, *J. Appl. Phys.* **85**(10), 6957–7022 (1999)
- 23.7 K. R. Elliott: *Appl. Phys. Lett.* **42**(3), 274–276 (1983)
- 23.8 K. Kurusu, Y. Suzuki, H. Takami: *J. Electrochem. Soc.* **136**, 1450–1452 (1989)
- 23.9 K. Terashima et al.: *Jpn. J. Appl. Phys.* **79**, 463–468 (1984)
- 23.10 L. R. Weisberg, F. D. Rosi, P. G. Herkart: *Properties of Elemental and Compound Semiconductors*, Vol. 5, ed. by H. C. Gatos (Interscience, New York 1960) pp. 25–67
- 23.11 J. M. Woodall: *Electrochem. Technol.* **2**, 167–169 (1964)
- 23.12 G. R. Cronin, R. W. Haisty: *J. Electrochem. Soc.* **111**, 874–877 (1964)
- 23.13 G. Martinez, A. M. Hennel, W. Szuszkiewicz, M. Balkanski, B. Clerjaud: *Phys. Rev. B* **23**, 3920 (1981)
- 23.14 T. P. Chen, T. S. Huang, L. J. Chen, Y. D. Gou: *J. Cryst. Growth* **106**, 367 (1990)
- 23.15 F. Moravec, B. Stepánek, P. Doubrava: *Cryst. Res. Technol.* **26**, 579–585 (1991)
- 23.16 E. P. A. Metz, R. C. Miller, R. Mazelsky: *J. Appl. Phys.* **33**, 2016 (1962)
- 23.17 J. B. Mullin, R. J. Heritage, C. H. Holliday, B. W. Straughan: *J. Cryst. Growth* **3–4**, 281–285 (1968)
- 23.18 J. R. Oliver, R. D. Fairman, R. T. Chen: *Electron. Lett.* **17**, 839–841 (1981)
- 23.19 H. M. Hobgood, L. B. Ta, A. Rohatgi, G. W. Eldridge, R. N. Thomas: Residual impurities and defect levels in semi-insulating GaAs grown by liquid encapsulated Czochalski, *Proc. Conf. Semi-insulating III–V Materials*, Evian, France 1982, ed. by S. Makram-Ebeid, B. Tuck (Shiva, Nantwich 1982) 30
- 23.20 D. E. Holmes, R. T. Chen, K. R. Elliott, C. G. Kirkpatrick: *IEEE Trans. Electron. Dev.* **29**, 1045 (1982)
- 23.21 D. E. Holmes, R. T. Chen, K. R. Elliott, C. G. Kirkpatrick: *Appl. Phys. Lett.* **40**, 46–48 (1982)
- 23.22 K. Laithwaite, R. C. Newman, J. F. Angress, G. A. Gledhill: *Inst. Phys. Conf. Ser.* **33**, 133 (1977)
- 23.23 A. S. Jordan, R. Caruso, A. R. von Neida: *Bell Syst. Tech. J.* **59**, 593 (1980)
- 23.24 M. R. Brozel, J. B. Clegg, R. C. Newman: *J. Phys. D* **11**, 1331 (1978)
- 23.25 P. J. Doering, B. Friedenreich, R. J. Tobin, P. J. Pearah, J. P. Tower, R. M. Ware: *Proc. 6th Conf. Semi-Insulating III–V Materials*, Toronto, Canada 1990, ed. by A. G. Milnes, C. I. Miner (IOP, London 1990) 173–181
- 23.26 U. Lambert, G. Nagel, H. Rufer, E. Tomzig: *Proc. 6th Conf. Semi-Insulating III–V Materials*, Toronto, Canada 1990, ed. by A. G. Milnes, C. I. Miner (IOP, London 1990) 183–188
- 23.27 J. M. Baranowski, P. Trautman: *Properties of Gallium Arsenide*, 3 edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 341–357
- 23.28 I. Grant, D. Rumsby, R. M. Ware, M. R. Brozel, B. Tuck: *Proc. Conf. on Semi-Insulating III–V Materials*, Evian, France 1982, ed. by S. Makram-Ebeid, B. Tuck (Shiva, Nantwich 1982) 98–106
- 23.29 M. S. Skolnick, M. R. Brozel, L. J. Reed, I. Grant, D. J. Stirland, R. M. Ware: *J. Electron. Mater.* **13**, 107–125 (1984)
- 23.30 A. G. Cullis, P. G. Augustus, D. J. Stirland: *J. Appl. Phys.* **51**, 2256 (1980)
- 23.31 D. Rumsby, I. Grant, M. R. Brozel, E. J. Foulkes, R. M. Ware: Electrical behaviour of annealed LEC GaAs, *Proc. Conf. on Semi-Insulating III–V Materials*, Kah-Nee-Ta, ed. by D. C. Look, J. S. Blakemore (Shiva, Nantwich 1984) 165–170
- 23.32 O. Oda: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 591–595
- 23.33 E. M. Monberg, H. Brown, C. E. Bormer: *J. Cryst. Growth* **94**, 643–650 (1989)
- 23.34 E. D. Bourret, E. C. Merk: *J. Cryst. Growth* **110**, 395–404 (1991)
- 23.35 M. R. Brozel, I. R. Grant: Growth of gallium arsenide. In: *Bulk Crystal Growth*, ed. by P. Capper (Wiley, Chichester 2005) pp. 43–71
- 23.36 M. S. Tyagi: *Introduction to Semiconductor Materials and Devices* (Wiley, New York 1991)
- 23.37 M. G. Astles: *Liquid Phase Epitaxial growth of III–V Semiconductor Materials and their Device Applications* (IOP, Bristol 1990)
- 23.38 K. Somogyi: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 625–638
- 23.39 M. Razeghi: *The MOCVD Challenge* (Hilger, Bristol 1989)
- 23.40 G. B. Stringfellow: *Organometallic Vapor Phase Epitaxy: Theory and Practice* (Academic, Boston 1989)
- 23.41 K. F. Jensen: *Handbook of Crystal Growth*, Vol. 3b, ed. by D. T. J. Hurle (Elsevier Science, Amsterdam 1994) p. 3541

- 23.42 D. W. Kisker: *Handbook of Crystal Growth*, Vol. 3b, ed. by D. T. J. Hurle (Elsevier Science, Amsterdam 1994) p. 393
- 23.43 G. B. Stringfellow: *Handbook of Crystal Growth*, Vol. 3b, ed. by D. T. J. Hurle (Elsevier Science, Amsterdam 1994) p. 349
- 23.44 L. Samuelson, W. Seifert: *Handbook of Crystal Growth*, Vol. 3b, ed. by D. T. J. Hurle (Elsevier Science, Amsterdam 1994) p. 745
- 23.45 M. A. Herman, H. Sitter: *Molecular Beam Epitaxy: Fundamentals and Current Status* (Springer, Berlin Heidelberg New York 1989)
- 23.46 K. R. Evans: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 655–662
- 23.47 F. W. Smith, A. R. Calawa, C. L. Chen, M. J. Manfra, L. J. Mahoney: *IEEE Electron. Dev. Lett.* **9**, 77 (1988)
- 23.48 M. Missous: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 679–683
- 23.49 J. F. Whitaker: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 693–701
- 23.50 N. X. Nguyen, U. K. Mishra: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 689–692
- 23.51 C. R. Abernathy: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 663–671
- 23.52 J. W. Matthews, A. E. Blakeslee: *J. Cryst. Growth* **27**, 118 (1974)
- 23.53 B. Tuck: *Introduction to Diffusion in Semiconductors* (Peregrinus, Stevenage 1974)
- 23.54 B. Tuck: *Atomic Diffusion in III–V Semiconductors* (A. Hilger, Bristol 1988)
- 23.55 K. A. Khadim, B. Tuck: *J. Mater. Sci.* **7**, 68–74 (1972)
- 23.56 J. Crank: *The Mathematics of Diffusion* (Clarendon, Oxford 1975)
- 23.57 K. P. Roenker: *Microelectron. Reliab.* **35**, 713 (1995)
- 23.58 S. A. Stockman: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 101–116
- 23.59 T. Ahlgren, J. Likonen, J. Slotte, J. Raisanen, M. Rajatorra: *Phys. Rev. B* **56**, 4597–4603 (1997)
- 23.60 M. R. Brozel, E. J. Foulkes, B. Tuck, N. K. Goswami, J. E. Whitehouse: *J. Phys. D* **16**, 1085–1092 (1983)
- 23.61 B. J. Sealy: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 765–782
- 23.62 M. S. Abrahams, C. J. Buicocchi: *J. Appl. Phys.* **36**, 2855 (1965)
- 23.63 R. T. Blunt, S. Clarke, D. J. Stirland: *IEEE Trans. Electron. Dev.* **29**, 1039 (1982)
- 23.64 A. R. Lang: Recent Applications of X-Ray Topography. In: *Modern Diffraction and Imaging Technique in Materials Science*, ed. by S. Amelinckx, G. Gevers, J. Van Landuyt (North Holland, Amsterdam 1978) pp. 407–479
- 23.65 G. T. Brown, C. A. Warwick: *J. Electrochem. Soc.* **133**, 2576 (1986)
- 23.66 P. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan: *Electron Microscopy of Thin Crystals* (Krieger, Malabar 1977)
- 23.67 M. R. Brozel: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) p. 377
- 23.68 L. Breivik, M. R. Brozel, D. J. Stirland, S. Tuzemen: *Semicond. Sci. Technol.* **7**, A269–A274 (1992)
- 23.69 G. Jacob: Proc. Conf. on Semi-Insulating III–V Materials, Evian, France 1982, ed. by S. Makram-Ebeid, B. Tuck (Shiva, Nantwich 1982) 2
- 23.70 T. Ogawa, T. Kojima: *Mater. Sci. Monogr.* **44**, 207–214 (1987)
- 23.71 R. C. Newman: *Infra-Red Studies of Crystal Defects* (Taylor Francis, London 1973)
- 23.72 R. Murray: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 227–234
- 23.73 R. J. Wagner, J. J. Krebs, G. H. Strauss, A. M. White: *Soild State Commun.* **36**, 15 (1980)
- 23.74 G. M. Martin: *Appl. Phys. Lett.* **39**, 747 (1981)
- 23.75 M. Skowronski, J. Lagowski, H. C. Gatos: *J. Appl. Phys.* **59**, 2451 (1986)
- 23.76 J. Dabrowski, M. Scheffler: *Phys. Rev. Lett.* **60**, 2183 (1988)
- 23.77 T. J. Chadi, K. J. Chang: *Phys. Rev. Lett.* **60**, 2187 (1988)
- 23.78 M. R. Brozel, I. Grant, R. M. Ware, D. J. Stirland: *Appl. Phys. Lett.* **42**, 610–12 (1983)
- 23.79 M. S. Skolnick, L. J. Reed, A. D. Pitt: *Appl. Phys. Lett.* **44**, 447–449 (1984)
- 23.80 M. R. Brozel, M. S. Skolnick: Near band edge “Reverse Contrast” images in GaAs, Proc. Conf. Semi-Insulating III–V Materials, Hakone 1986, ed. by H. Kukimoto, S. Miyazawa (Shiva, Nantwich 1986) 109
- 23.81 C. Le Berre et al.: *Appl. Phys. Lett.* **66**, 2534 (1995)
- 23.82 M. R. Brozel, S. Tuzemen: *Mater. Sci. Eng. (B)* **28**, 130–133 (1994)
- 23.83 G. M. Martin, A. Mitonneau, A. Mircea: *Electron. Lett.* **13**, 191–193 (1977)
- 23.84 A. Mitonneau, G. M. Martin, A. Mircea: *Electron. Lett.* **13**, 666–668 (1977)
- 23.85 A. Mitonneau, A. Mircea, G. M. Martin, D. Pons: *Revue de Phys. Appl.* **14**, 853–861 (1979)
- 23.86 S. A. Stockman, A. W. Hanson, S. L. Jackson, J. E. Baker, G. E. Stillman: *Appl. Phys. Letts.* **62**, 1248 (1992)
- 23.87 N. Watanabe, T. Nittono, H. Ito: *J. Cryst. Growth* **145**, 929 (1994)
- 23.88 M. R. Melloch, N. Otsuka, J. M. Woodall, A. C. Warren, J. L. Freeouf: *Appl. Phys. Lett.* **57**, 1631 (1990)
- 23.89 Z. Liliental-Weber, A. Claverie, J. Washburn, F. Smith, A. R. Calawa: *Appl. Phys. A* **53**, 141 (1991)

- 23.90 A. C. Warren, J. M. Woodall, J. L. Freeouf, D. Grischkowsky, D. T. Melloch: Appl. Phys. Lett. **57**, 1331–1333 (1990)
- 23.91 D. E. Bliss, W. Walukiewicz, J. W. Ager III, E. E. Haller, K. T. Chan, S. J. Tamigawa: Appl. Phys. **71**, 1699 (1992)
- 23.92 D. J. Keeble, M. T. Umlor, P. Asoka-Kumar, K. G. Lynn, P. W. Cooke: Appl. Phys. Lett. **63**, 87 (1993)
- 23.93 D. K. Schroder: *Semiconductor Material and Device Characterization*, 2nd edn. (Wiley Interscience, New York 1998)
- 23.94 Ch. H. Alt: Appl. Phys. Lett. **54**, 1445 (1989)
- 23.95 D. C. Look: *Electrical Characterization of GaAs materials and Devices*, Design Meas. Electron. Eng. Ser. 1989 (Wiley, Chichester 1989)
- 23.96 R. Stibal, J. Windscheif, W. Jantz: Semicond. Sci. Technol. **6**, 995–1001 (1991)
- 23.97 M. Wickert, R. Stibal, P. Hiesinger, W. Jantz, J. Wagner: High resolution EL2 and resistivity topography of Si GaAs wafers, Proc. SIMC-X, Berkeley, CA 1998, ed. by Z. Liliental-Weber, C. Miner (IEEE, 1999) 21–24
- 23.98 G. M. Martin, J. P. Farges, G. Jacob, J. P. Hal-lais, G. Poiblaud: J. Appl. Phys. **51**, 2840–2852 (1980)
- 23.99 G. E. Stillman, C. M. Wolfe, J. O. Dimmock: *Semicond. Semimet.*, Vol. 21 (Academic, New York 1977) p. 169
- 23.100 M. N. Afsar, K. J. Button, G. L. McCoy: Inst. Phys. Conf. Ser. **56**, 547–555 (1980)
- 23.101 P. Blood, J. W. Orton: *The Electrical Characterization of Semiconductors: Majority Carriers and Electron States* (Academic, London 1992)
- 23.102 D. V. Lang, L. C. Kimmerling: IOP Conf. Ser. **23**, 581 (1975)
- 23.103 D. V. Lang: J. Appl. Phys. **45**, 3023 (1974)
- 23.104 L. Dobaczewski, P. Kaczor, I. D. Hawkins, A. R. Peaker: J. Appl. Phys. **76**, 194 (1994)
- 23.105 H. B. Bebb, E. W. Williams: Photoluminescence I: Theory. In: *Semicond. Semimet.*, Vol. 8, ed. by R. K. Willardson, A. C. Beer (Academic, New York 1972) pp. 181–320
- 23.106 P. J. Dean: Prog. Cryst. Growth Charact. **5**, 89–174 (1982)
- 23.107 M. Tajima, T. Iino: Jpn. J. Appl. Phys. **28**, L841–844 (1989)
- 23.108 C. J. Miner, C. J. L. Moore: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 320–332
- 23.109 O. Oda, H. Yamamoto, M. Seiwa, G. Kano, T. Inoue, M. Mori, H. Shimakura, M. Oyake: Semicond. Sci. Technol. **7**, A215 (1992)
- 23.110 B. J. Skromme, C. J. Sandroff, E. Yablonovitch, T. Gmitter: Appl. Phys. Lett. **51**, 24 (1987)
- 23.111 C. J. Miner: Semicond. Sci. Technol. **7**, A10 (1992)
- 23.112 A. M. Huber, C. Grattapain: SIMS Analysis of III–V compound microelectronic materials. In: *Analysis of Microelectronic Materials and Devices*, ed. by M. Grasserbauer, H. W. Werner (Wiley, New York 1991) p. 305
- 23.113 B. K. Tanner, D. K. Bowen: *Characterization of Crystal Growth Defects by X-Ray Methods* (Plenum, New York 1980)
- 23.114 B. K. Tanner: *X-Ray Topography and Precision Diffractometry of Semiconductor Materials*, ed. by T. J. Shaffner, D. K. Schroder (Electrochem. Soc., Pennington 1988) pp. 133–149
- 23.115 S. P. Kwok: J. Vac. Sci. Tech. B **4**, 6 (1986)
- 23.116 B. K. Ridley, T. B. Watkins: Proc. Phys. Soc. **78**, 293 (1961)
- 23.117 C. Hilsum: Proc. IRE **50**, 185 (1962)
- 23.118 J. B. Gunn: Solid State Commun. **1**, 88 (1963)
- 23.119 M. P. Shaw: *The Physics and Instabilities of Solid State Electron Devices* (Kluwer Academic/Plenum, Dordrecht 1992) pp. 830–835
- 23.120 C. G. Discus et al.: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996)
- 23.121 J. Singh: *Physics of Semiconductors and Their Heterostructures* (McGraw–Hill, New York 1994)
- 23.122 Y. Chang, F. Kai: *GaAs High-Speed Devices* (Wiley, New York 1994)
- 23.123 J. M. Golio: *Microwave Metal Semiconductor Field Effect Transistors and High Electron Mobility Transistors* (Artech House, London 1991)
- 23.124 W. Liu: *Fundamentals of III–V Devices: HBTs, MESFETs and HFETs/HEMTs* (Wiley, New York 1999)
- 23.125 R. H. Wallis: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 811–819
- 23.126 W. Shockley: US Patent 2 569 347 (1951)
- 23.127 H. Kroemer: Proc. IRE **45**, 1535 (1957)
- 23.128 E. F. Schubert: *Light Emitting Diodes* (Cambridge Univ. Press, Cambridge 2003)
- 23.129 E. F. Schubert: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 874–886
- 23.130 E. Kapon (ed): *Semiconductor Lasers I. Fundamentals* (Academic, New York 1998)
- 23.131 T. Numai: *Fundamentals of Semiconductor Lasers* (Springer, Berlin Heidelberg New York 2004)
- 23.132 L. A. Coldren, S. W. Corzine: *Diode Lasers and Photonic Integrated Circuits* (Wiley Interscience, New York 1995)
- 23.133 R. M. Kolbas: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 887–905
- 23.134 G. Morthier, P. Vankwikelberge: *Handbook of Distributed Feedback Laser Diodes* (Artech House, London 1997)

- 23.135 P. K. Bhattacharya: *Properties of Gallium Arsenide*, 3rd edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 861–873
- 23.136 T. E. Sale: *Vertical Cavity Surface Emitting Lasers*, Electron. El. Res. Stud. Optoelectronics 5 (Research Studies, 2003)
- 23.137 Y. J. Yang: Appl. Phys. Lett. **62**, 600–602 (1993)
- 23.138 L. A. Coldren, S. W. Corzine: *Diode Lasers and Photonic Integrated Circuits* (Wiley, New York 1995)
- 23.139 H. Yamamoto, M. Asada, Y. Suematsu: Electron. Lett. **21**, 579 (1985)
- 23.140 J. M. Woodall, H. J. Hovel: Appl. Phys. Lett. **30**, 492 (1977)
- 23.141 L. J. Kozlowski et al.: IEEE Trans. Electron. Dev. **38**, 1124 (1991)
- 23.142 C. M. Buttar: GaAs detectors and related compounds, Nucl. Inst. Phys. Res. A **395**, 1–8 (1997)
- 23.143 L. P. Sadwick, R. J. Hwu: *Properties of Gallium Arsenide*, 3 edn., ed. by M. R. Brozel, G. E. Stillman (INSPEC, London 1996) pp. 948–962

24. High-Temperature Electronic Materials: Silicon Carbide and Diamond

The physical and chemical properties of wide-band-gap semiconductors make these materials an ideal choice for device fabrication for applications in many different areas, e.g. light emitters, high-temperature and high-power electronics, high-power microwave devices, micro-electromechanical system (MEM) technology, and substrates for semiconductor preparation. These semiconductors have been recognized for several decades as being suitable for these applications, but until recently the low material quality has not allowed the fabrication of high-quality devices. In this chapter, we review the wide-band-gap semiconductors, silicon carbide and diamond.

Silicon carbide electronics is advancing from the research stage to commercial production. The commercial availability of single-crystal SiC substrates during the early 1990s gave rise to intense activity in the development of silicon carbide devices. The commercialization started with the release of blue light-emitting diode (LED). The recent release of high-power Schottky diodes

24.1	Material Properties and Preparation	540
24.1.1	Silicon Carbide	540
24.1.2	Diamond	544
24.2	Electronic Devices	547
24.2.1	Silicon Carbide	547
24.2.2	Diamond	551
24.3	Summary	557
	References	558

was a further demonstration of the progress made towards defect-free SiC substrates.

Diamond has superior physical and chemical properties. Silicon-carbide- and diamond-based electronics are at different stages of development. The preparation of high-quality single-crystal substrates of wafer size has allowed recent significant progress in the fabrication of several types of devices, and the development has reached many important milestones. However, high-temperature studies are still scarce, and diamond-based electronics is still in its infancy.

The electronic revolution of the 20th century is mainly based on silicon, which can be regarded as the first-generation semiconductor. Around the turn of the 21st century gallium arsenide and indium phosphide have evolved as second-generation semiconductors, constituting the base for the wireless and information revolution. Now at the start of the 21st century, the wide-band-gap semiconductors silicon carbide and gallium nitride are on the rise and may be regarded as third-generation semiconductors used in the electronic and optoelectronic industries. Moreover given diamond's superior properties and the recent surge of research on diamond preparation and fabrication of diamond-based electronic devices, one might speculate that diamond may be a future-generation semiconductor.

The effects of temperature on materials and devices have been of great interest throughout the history of

semiconductor research. The aim has been to investigate the high-temperature limits of materials and to enhance high-temperature semiconductor device performance. The development of semiconductor devices for reliable operation for an extended period at high temperatures is a complex process in which a number of physical effects connected with increasing temperature [24.1, 2] have to be considered. The term high temperature is not defined in a unique way in the literature and has a different meaning depending on the semiconductor under consideration and the area of application of semiconductor devices. The definition of high temperature often cited in the literature is temperatures above 125 °C [24.2, 3], since 125 °C is frequently specified as the upper limit at which standard commercial silicon devices function properly, although tests on standard commercial components indicate that even 150 °C may be applicable for selected silicon components [24.3].

Silicon is still the dominant semiconductor and silicon devices are still being developed. The most common and cost-effective integrated circuit technology is now silicon complementary metal–oxide–semiconductor (CMOS), which is able to operate up to 200 °C. The silicon-on-insulator (SOI) technology extended the operational temperature of CMOS circuits to 300 °C [24.4–8]. In addition, devices based on gallium arsenide and related alloys, which are commercialized to a lesser degree than silicon, are also candidates for high-temperature operation beyond 300 °C. The short-term operation of GaAs devices at temperatures as high as 500 °C has been reported [24.9, 10].

A survey of the literature indicates that 300 °C can be regarded as a dividing point from several standpoints, e.g. packaging, wiring, connecting, etc. [24.1, 2, 4]. This temperature is approximately the maximum temperature at which low-power silicon or conventional gallium arsenide devices can function reliably. The intrinsic carrier concentration for several semiconductors as a function of temperature is shown in Fig. 24.1. The control of the free-carrier concentration is vital for the performance of all semiconductor devices. The intrinsic carrier concentration (n_i) is exponentially dependent on the temperature:

$$n_i = \sqrt{N_C N_V} e^{-E_g/2k_B T}, \quad (24.1)$$

where E_g is the band gap, k_B is Boltzmann's constant and T is the temperature in Kelvin. Evidently, at temperatures above 300 °C, SiC, GaN and diamond and AlN have much lower intrinsic carrier concentrations than Si

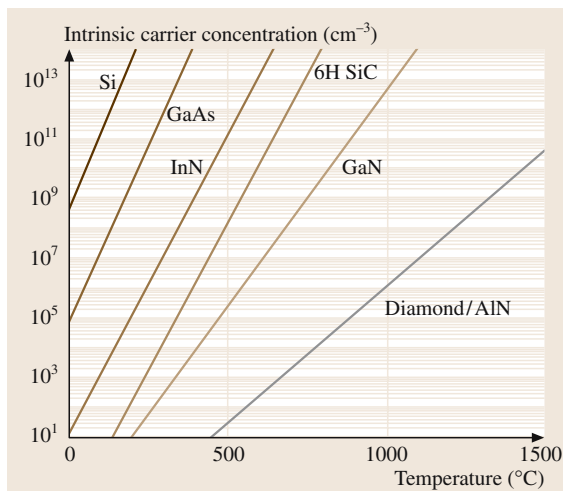


Fig. 24.1 Intrinsic carrier concentration as a function of temperature of several semiconductors. (After [24.9])

and GaAs. This implies that devices designed for higher temperatures should be fabricated from wide-band-gap semiconductors to avoid the deteriorating effects of thermally generated carriers.

The wide-band-gap third-generation semiconductors, SiC and GaN (including the III–nitride systems e.g. AlGaN), have been recognized for over three decades as materials which are well suited for high-temperature electronics and for light emitters, but until recently, the low material quality has not allowed the production of high-quality devices. The availability of single-crystal SiC wafers at the start of the 1990s initiated a great deal of activity towards the development of SiC-based devices, and their commercialization started with the release of blue light-emitting diode (LED). The availability of commercial high-quality substrates meant that more research has been carried out on SiC than on the GaN and III–nitride systems. The SiC devices have the advantages of a more mature semiconductor material growth and device fabrication technology. Furthermore, GaN and III–nitride crystals have mostly been grown by heteroepitaxy on e.g. sapphire and SiC, since a viable GaN substrate technology does not exist. Unfortunately, GaN crystals always contain more defects than SiC and the current aim is to reduce the surface defect densities in GaN from current densities of the order 10⁸ cm⁻² to 10⁵ cm⁻². The unavailability of low-defect-density substrates and defect-free material limits the ability to fabricate high-quality GaN devices. The discussion of wide-band-gap semiconductors must mention AlN since it has one of the largest band gaps (wurtzite: 6.23 eV, and zincblende: 6.0 eV [24.11]). The growth of defect-free AlN crystals (as with the GaN) is an outstanding issue.

The reduction of the defect density and the effects of specific defects of third-generation semiconductors are the most urgent current problems that must be solved. Diamond is a future-generation semiconductor which is at a different stage of research than the third-generation semiconductors, particularly SiC, which is far more developed than diamond.

The research in wide-band-gap semiconductors has been driven by the need for light emitters, high-temperature and high-power industrial applications, and microwave power applications. A variety of applications e.g. in aircraft and space systems, automotive electronics, deep-well drilling, energy production centers etc., would benefit from power devices that function at high temperatures [24.12–16]. When the ambient temperature is too high, the performance-enhancing electronics presently used beneficially to monitor and control crucial hot sections must reside in cooler areas, this is achieved

by their remote location or actively by cooling with air or liquids. These thermal management approaches introduce additional overheads that can have a negative impact relative to the desired benefits when considering the overall system performance. The additional overhead, in the form of longer wires, more connectors and plumbing for the cooling system, can add undesired size and weight to the system, and increased complexity that corresponds to an increased potential for failure. The economic benefits of high-temperature electronics for various systems are likely to be orders of magnitude greater than the total market for actual high-temperature electronics. The world market for high-temperature electronics between 2003–2008 is predicted to increase from 400 to 900 million US dollars, which is substantially lower than the world's total semiconductor electronic market [24.1]. The situation can be dramatically described as follows, a mere handful of high-temperature electronic chips that may cost a few hundred dollars, can optimize the performance of a very large number of systems, thus saving many millions of dollars, e.g. deep-well drilling [24.5].

A survey of the potential industrial users of high-temperature electronics revealed that the majority of applications for high-temperature electronics operate in the range 150–300 °C [24.1, 2, 4]. The recent development of silicon and gallium arsenide electronics and their cost (silicon technology is much cheaper than SiC), indicates that wide-band-gap semiconductor devices are unlikely in the near future to be used in low-power electronics applications for temperatures up to 300 °C. These devices maybe used for application which cannot be satisfied by available technologies such as SOI, and for temperatures above 300 °C. However, in order to realize viable low-power SiC devices for the temperature range 300–600 °C, long-term reliability of electronic circuits must be achieved [24.1].

The performances of silicon power devices have almost reached their theoretical limits [24.17]. The practical operation of Si power devices at ambient temperatures higher than 200 °C appears problematic, as self-heating due to current flow at higher power levels results in high internal junction temperatures and leakage. The overall goal for high-temperature power-electronic circuits is to reduce power losses, volume, weight, and at least the costs of the system. The continuous progress in high-temperature electronics creates a demand for unique material properties, novel processing technologies and electronic devices. The physical and chemical properties required for meeting the demands of the high-temperature and high-power applications can only be

found in wide-band-gap semiconductors, which offer a number of advantages over corresponding devices fabricated from silicon. These include higher temperature stability, higher chemical stability, higher thermal conductivity, and higher breakdown field. Various device implementations not only use these standard semiconductor parameters, but also the special peculiarities these materials exhibit, e.g. aluminium nitride and gallium nitride, unlike diamond and silicon carbide, have a direct band gap and have complete miscibility with each other and with indium nitride. This is important for the implementation of optoelectronic device since it allows the band gap to be controlled, and thus the wavelength of the spectral characteristic maximum [24.4]. The wide-band-gap silicon carbide and diamond are next discussed in this review.

The properties of silicon carbides make it an excellent material for high-power devices operating at temperatures up to 600 °C and above, and at frequencies around 20 GHz. Within power electronics, SiC has the potential to replace Si-based diodes and insulated gate bipolar transistors (IGBTs), and Si gate turn-off (GTO) thyristors, which are part of the mass market of discrete power devices in general and in converter systems in particular. The power losses in SiC switches are two orders of magnitude lower compared with Si devices, thus SiC devices have a large potential for applications in e.g. uninterrupted power systems (UPS), motor controls, etc. The maximum operating temperature of a Schottky diode in SiC may be limited by an increasing leakage currents, but active power devices for operation at high temperature has been presented. U-shaped-trench metal–oxide–semiconductor field-effect transistors (UMOSFET) made from SiC that operate up to 450 °C and thyristors (6 A, 700 V) that operate at 350 °C have been presented. Furthermore, SiC MOSFETs have been reported to operated even at 650 °C, and devices based on n-type-channel metal–oxide–semiconductor (NMOS) technology, which is an integrated operational amplifier, have been reported to work at 300 °C [24.18–22]. The properties and preparation of SiC are elucidated in the next section.

Among the wide-band-gap semiconductors, diamond has the best physical, chemical and electrical properties [24.23], unmatched by any other material. The properties of interest relevant to high-temperature high-frequency power electronics are the large band-gap energy (5.5 eV), the breakdown electric field (10 MV/cm), the carrier mobilities (≈ 2200 and ≈ 1600 cm²/V s for electrons and holes, respectively), the thermal conductivity (10–20 W/cmK), the low dielectric constant (5.5),

and the excellent resistance to radiation. Diamond can be found naturally or must be synthesized. In nature diamond occurs as single crystals only, whereas synthetic diamond can be prepared as single crystals, or as a polycrystalline or as a nanocrystalline material.

The discovery that diamond can be grown by the chemical vapor deposition (CVD) technique has opened up some of the expected applications of diamond. However, the utilization of diamond's many unique properties in electronics has so far been limited among others by the unavailability of large-area high-quality diamond and that only p-type (acceptor-type impurity) diamond with high hole densities are available today. The n-type (donor-type impurity) diamond with high electron densities would find many applications, apart from the fundamental interest to realize pn-junctions and other electronic devices in diamond. The n-type diamond is expected to be a better electron emitter for field emission, photo emission, and ion or electron impact-induced emission, and may also serve as a better inert electrode for electrochemical applications.

Nevertheless, many studies have been reported with natural, high-pressure high-temperature (HPHT) synthesized and polycrystalline CVD diamonds [24.24, 25]. The pn-junctions were formed from boron- and phosphorus-doped diamond films, and from boron- and nitrogen-doped diamond films, respectively. The diamond films with high crystalline perfection were grown epitaxially on diamond single crystals. The current–

voltage (I – V) characteristic of the boron/nitrogen pn-junction diode was studied up to 400 °C. The combination of two boron/nitrogen pn-junctions, a bipolar junction transistor (BJT) which can operate in direct-current (DC) mode up to 200 °C was fabricated. The fabrication of many types of field-effect transistors (FETs) for both DC and radio-frequency (RF) modes has crossed many important milestones. The cutoff frequency of 1.7 GHz and a maximum drain current of 360 mA/mm were measured for a metal–semiconductor field-effect transistor (MESFET) with a gate length of 0.2 μm . Recently, a FET functioning up to 81 GHz was reported by a collaboration between Nippon Telegraph and Telephone Corp. and the University of Ulm in Germany. The research groups fabricated T-shaped gates on a diamond layer with a carrier mobility of 130 cm^2/Vs [24.26]. In addition, Schottky diodes that function up to 1000 °C were fabricated from either single-crystal or polycrystalline diamond. Low-resistance thermostable resistors deposited on ceramic substrates have been investigated for temperatures up to 800 °C. The temperature dependence of the field emission of nitrogen-doped diamond films has been investigated at temperatures up to 950 °C.

There has been much progress in the fabrication of diamond-based electronic devices and several types of devices have reached an important stage in their development. However, despite these developments, diamond-based electronics is still in its infancy.

24.1 Material Properties and Preparation

24.1.1 Silicon Carbide

The properties of silicon carbide makes it an excellent material for devices operating at high temperatures (600 °C and higher), high power (4H-SiC transistor: presently RF output power on the order of 5 W/mm), and high frequency [RF through X band (5.2–10.9 GHz) potentially to K band (20–40 GHz)]. The large band gap of silicon carbides (2.2, 3.26 and 3.0 eV for 3C-SiC, 4H-SiC and 6H-SiC, respectively) compared to the band gap of silicon (1.1 eV) enables devices to function at temperatures beyond 600 °C. The very high breakdown electric field of these materials (\approx 1.8, 3.5 and 3.8 MV/cm for 3C-SiC, 4H-SiC and 6H-SiC, respectively) which are approximately 10 times higher than that of Si (0.3 MV/cm), allows a reduction of the thickness of the conduction regions (for constant doping), which

results in very low specific conduction resistance. The 4H-SiC junctions exhibit a negative temperature coefficient, with a breakdown voltage that decreases by about 8% within the temperature range from room temperature to 623 °C [24.27]. The high thermal conductivity (\approx 4–4.5 W/cmK) permits a power density increase which facilitates a more compact or much higher power per area. The high saturation velocity of all three types of silicon carbide is high (\approx 2×10^7 cm/s) compared to the value for silicon (1×10^7 cm/s). The low carrier mobilities of silicon carbide is a disadvantage which limits RF performance at frequencies above the X band. The electron mobilities are of the order 900, 500 and 200 cm^2/Vs for 3C-SiC, 4H-SiC and 6H-SiC, respectively. The hole mobilities are of the order of 50 cm^2/Vs for all three types of SiC (for Si: \approx 1350 and \approx 500 cm^2/Vs for electrons and holes, respectively). The carrier mobilities

Table 24.1 Approximate values of physical properties for some semiconductors

Name	Bandgap E_g (eV)	Maximum electric field (V/cm)	Dielectric constant ϵ_s	Thermal conductivity (W/cmK)	Carrier mobility (cm^2/Vs)
Si	1.1	3×10^5	11.8	1.5	1350 480
GaAs	1.4	3.5×10^5	10.9	0.8	8600 250
SiC	3.3	2.5×10^6	9.8	4.9	980 200
GaN	3.4	2×10^6	7.8	1.4	2000
Diamond	5.5	1×10^7	5.5	10–20	1800 1600

of SiC are adequate however, for high-power devices in the X band. The properties of silicon carbide and diamond relevant for electronics are given in Table 24.1.

Noteworthy, is that silicon carbide has a close lattice match with III–nitrides, which makes it a preferred substrate material for nitride-based electronic and optoelectronic devices. The commercial production of large substrates which have improved electronic and optoelectronic properties constitutes a milestone in their application. These materials have been used (among

others) for large-scale production of green, blue and ultraviolet light-emitting diodes. Unfortunately, unavailability of high-quality defect-free SiC substrates is slowing down the pace of transition from research and development (R&D) to production of SiC devices, which may include high-power solid-state switches or diodes for electrical power control, and high-power-density microwave transistors.

Silicon carbide occurs in a large number of polytype structures. The number of polytypes in the literature

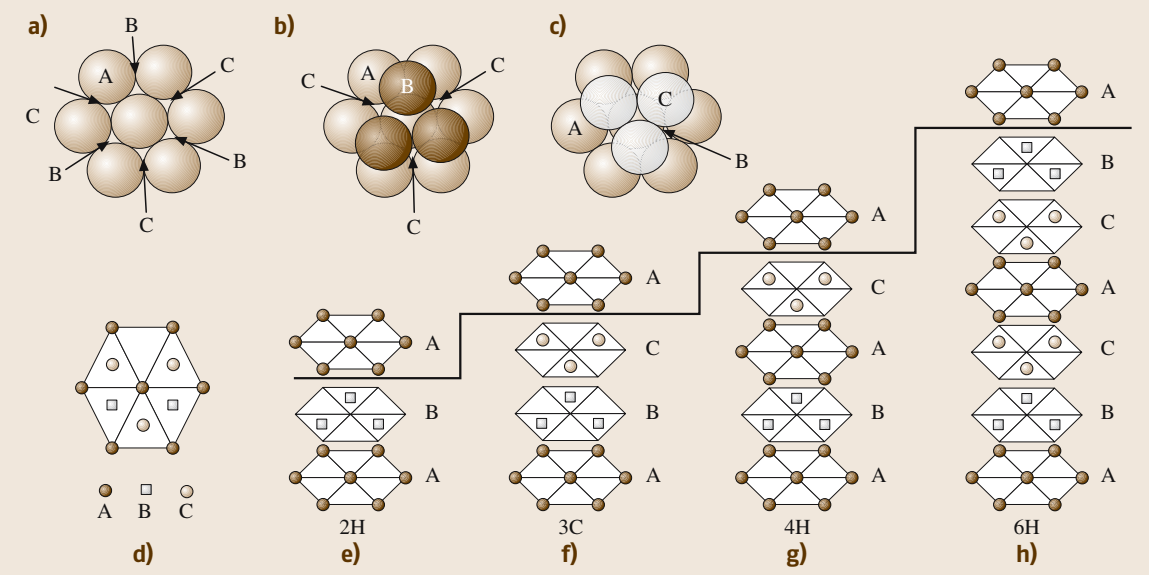


Fig. 24.2a–d The stacking sequence of double layers of the four most common SiC polytypes, (a) locations of C atoms, labeled A, in the first biatomic layer in the {0001} plane; (b), (c) optional positions of C atoms, labeled B and C, respectively, in the next biatomic layer above the first layer; (d) first biatomic layer with six C atoms and the optional positions of the three C atoms in the next biatomic layer, (e)–(h) stacking sequence of the most common SiC polytypes. The solid line indicates the completion of the unit cell in the [0001] direction (After [24.28])

varies between 150 and 250. These polytypes are differentiated by the stacking sequence of the biatomic close-packed layers. A detailed study of silicon carbide's polytypism was done in [24.29]. The most famous polytypes are the hexagonal 4H and 6H, cubic 3C, and rhombohedral 15R structures. Not all types are easy to grow; only 4H and 6H polytypes are available as substrate materials. In a single bilayer of SiC each C atom is tetrahedrally bonded to four Si atoms – to three ones within the layer and to one in the next layer. Looking at a bilayer from the top in the direction of the *c*-axis [0001], the C atoms form a hexagonal structure, as shown in Fig. 24.2a and Fig. 24.2d. These are labeled 'A'. The C atoms of the next biatomic layer have the option to be positioned at the lattice sites 'B' or 'C', as shown in Fig. 24.2b and Fig. 24.2c, respectively. This is the stacking sequence defining a polytype. Figure 24.2d shows schematically the first bilayer, with six C atoms forming the hexagonal structure and the optional positions for the three C atoms in the next layer beyond the first layer. Figures 24.2e–h shows the stacking sequence for the most common SiC polytypes. The change in sequence has an impact on the properties of the ma-

terial, for example the band gap changes from 3.4 eV for 2H polytype to 2.4 eV for 3C polytype. Figures 24.3a–e shows the structure of the most common SiC polytypes viewed in the $\{11\bar{2}0\}$ plane, i. e. in the [0001] direction.

The leading manufacturer of substrates is Cree Inc., though new manufacturers have recently appeared. The crystalline quality in terms of a low defect density, and a specially low micropipe density the substrates of Cree Inc. are still ahead. The preparation of silicon carbide is complicated by the fact that it does not melt, it sublimates at temperature above 2000 °C, thus standard growth techniques, e.g. the Czochralski process by which large single-crystal ingots are produced by pulling a seed crystal from the melt, cannot be used. Silicon carbide crystals are grown by a sublimation method first developed by Lely in 1955 [24.30], and later extended to a seed sublimation technique by Tairov and Tsvetkov in 1978 [24.31]. This method is also termed physical vapor transport (PVD) growth. The crystals are grown by SiC deposition derived from Si and C molecular species provided by a subliming source of SiC placed in close proximity to the seed wafer.

A new high-temperature CVD (HTCVD) technique was developed in 1999 [24.28], where the growth rate can be tuned in such a way that a high-quality thick epitaxial layer with precisely controlled doping levels can be grown in a few hours, which is fast compared to conventional CVD, which takes days to grow a similar structure. As an alternative to CVD, sublimation epitaxy has also been demonstrated for the growth of thick epitaxial layers. The growth rate in sublimation epitaxy is also very high, of the order of a few hundred micrometers an hour. In devices where the control of the doping level is an important issue, sublimation epitaxy has been shown to work successfully.

The commercial availability of SiC substrates with increasing diameter and quality has been a prerequisite for advances in SiC device technology. The SiC substrates are available in two different polytypes, namely 6H- and 4H-SiC. The latter is relevant for electronic application due to its higher carrier mobility and wider band gap than 6H-SiC. SiC substrates have been in the market for over a decade, but the absence of defect-free growth is slowing the pace of transition from research and development to the production of power devices such as high-power solid-state switches or diodes for electrical power control and high-power-density microwave transistors. Three-inch 4H-SiC substrates have been commercially available since 2001, but it was only recently that their defect concentration have been reduced to levels that allows for the fabrication of com-

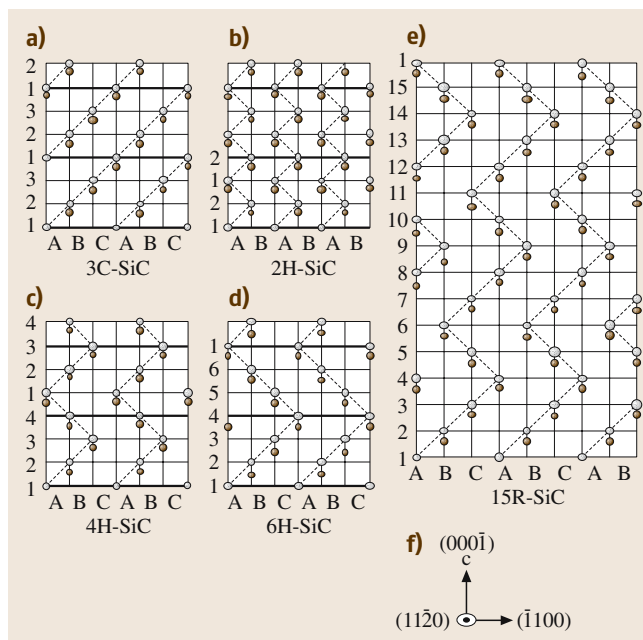


Fig. 24.3a–f Schematic representation of the structure of the most common SiC polytypes viewed in the $\{11\bar{2}0\}$ plane. The black dots represent C atoms and the open circles represent Si atoms. The bold line indicates the completion of the unit cell in the [0001] direction (After [24.28])

mercially viable high power switches. During 2003 Infineon and Cree Inc. released 10 Amps devices, which is clear evidence that substrates have reached an acceptable level of quality.

The important defects of SiC are different types of dislocations. The open core screw dislocations called micropipes are of particular concern for SiC due to their detrimental effects on power devices. Micropipes cause diodes to fail for voltages that are much smaller than the voltage at which avalanche breakdown occurs [24.33]. Progress in the development of the physical vapor deposition (PVD) technique during the last four years, has resulted in a significant reduction of the micropipe density in 3-inch 4H-SiC wafers, from a previously typical value above 100 cm^{-2} to a value as low as 0.22 cm^{-2} in R&D samples. The micropipe densities in commercially available substrates are $< 30\text{ cm}^{-2}$ and $< 80\text{ cm}^{-2}$ for n-type and semi-insulating materials, respectively. 100-mm 4H-SiC wafers are now under development. The micropipe densities for such wafers are $\approx 22\text{ cm}^{-2}$ and $\approx 55\text{ cm}^{-2}$ for n-type and semi-insulating crystals, respectively [24.34, 35].

Low-angle grain boundaries, also known as domain walls, are another class of defect that has to be reduced, since they are associated with leakage currents and failure in devices, and may cause wafers to crack during epitaxial processing. This type of defect can be observed through whole-wafer X-ray topography, though it is difficult to obtain quantifiable numbers on wafer quality by such measurements. This defect seems to be intimately related to the growth method used [24.36]. A class of defects known as threading screw dislocations are suggested to have an impact on the leakage behavior of Schottky diodes. The evidence suggests that these defects are introduced at the seed/growth interface by seed subsurface damage. The application of seed treatment reduces the density of dislocations in a 3-inch 4H-SiC wafer from a value of the order $3 \times 10^4\text{ cm}^{-2}$ to $3 \times 10^3\text{ cm}^{-2}$ [24.35]. The so-called basal plane dislocation is another important and significant defect. It has been shown that PiN device structures are susceptible to severe degradation of the forward voltage characteristics due to the presence of these defects in the active layer of the device [24.37]. The presence of these dislocations increases the resistance of the active layer of the device. It is critical to reduce the density of these dislocations in epitaxial layers for stable device production. A level of basal plane dislocations in the substrate which may be acceptable in order to allow reasonable yields of PiN diodes, is on the order of 100 cm^{-2} . The average density of the basal plane dislocations in 3-inch 4H-SiC wafers

is $1.5 \times 10^3\text{ cm}^{-2}$ [24.35]. Recently it was reported that this type and even other types of dislocations and defects can be reduced by growing the material along the *a*-face direction [24.32]. Single crystals of SiC are usually grown by the method termed *c*-face growth, where crystals are grown along the $\langle 0001 \rangle$ *c*-axis direction using a seed of $\{0001\}$ substrate. In the new method, known as a repeated *a*-face growth process, single crystals are grown along the *a*-axis $[1\bar{1}20]$ or $[1\bar{1}00]$ (both axis are called the *a*-axis in the report) direction in several steps. The *a*-face growth process is shown in Fig. 24.4.

The electrical properties of SiC substrates are related to the purity of the as-grown crystals. The substrate purity is dominated by the presence of residual nitrogen and boron impurities. The level of these impurities is critical for the production of undoped high-purity semi-insulating substrates with uniform and stable semi-insulating properties. High-purity 3-inch and 100-mm

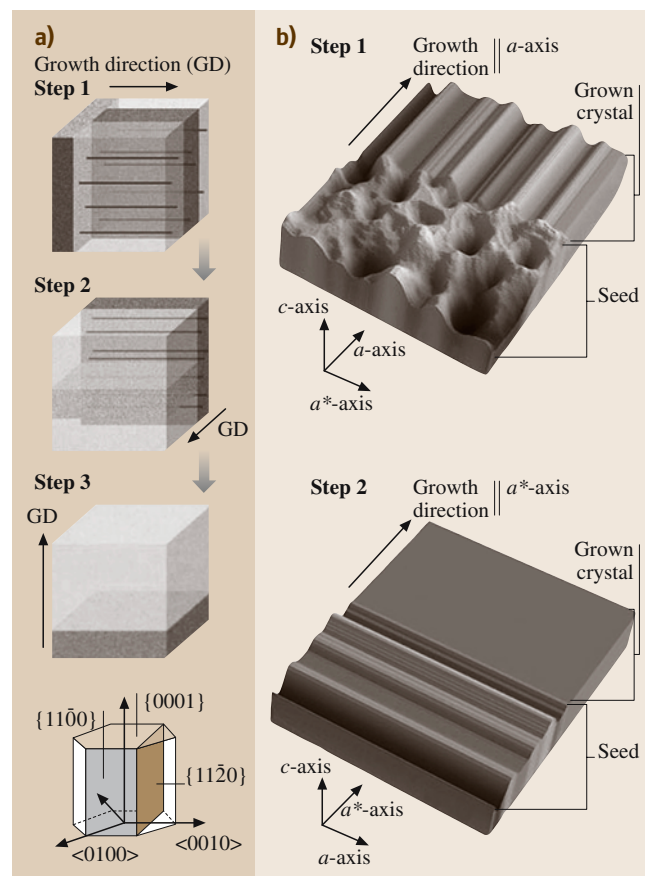


Fig. 24.4 Schematic picture of the *a*-face growth of SiC. (After [24.32])

4H-SiC substrates with low micropipe densities and uniform semi-insulating properties ($> 10^9 \Omega \text{ cm}$) over the full wafer diameter have been produced. These wafers had typical residual contamination densities $5 \times 10^{15} \text{ cm}^{-3}$ and $3 \times 10^{15} \text{ cm}^{-3}$ for nitrogen and boron, respectively [24.38].

Although most doping of SiC is obtained by an in situ method during epitaxial growth, additional selected-area doping is often required during fabrication of devices such as MOSFETs and lateral bipolar transistor. Due to the extremely low diffusion coefficient of dopant atoms in SiC even at very high temperatures ($\approx 2000^\circ\text{C}$), ion implantation is the only viable doping technique during device fabrication. The critical parameters of ion implantation of dopants in SiC are the temperature of SiC during implantation (from room temperature up to 900°C), as well as the subsequent annealing required to activate the dopant, performed at around 1700°C . Nitrogen is typically used as the n-type dopant, while Al is often the p-type dopant widely used during epitaxy. The element boron is a lighter element than Al and subsequently causes less lattice damage during implantation, and may eventually replace Al. However, a small amount of B also diffuses into the lightly doped drift-layer side during the annealing process, thereby degrading the junction.

The high bond strength of SiC means that room-temperature wet etches for this material do not exist, and so reactive-ion etching (RIE) is the standard method used. Frequently fluorine-based chemistries are used in which the silicon forms a volatile SiF_4 molecule and C is removed either as CO_2 or CF_4 . However, RIE is not regarded as a limitation since, as feature sizes decrease, dry etching processes are actually preferred to wet etching.

A unique advantage of SiC compared to other wide-band-gap semiconductors, is its ability to oxidize and form SiO_2 exactly as in Si technology. The oxidation rates are much lower for SiC than for Si, and are very dependent on if a silicon- or carbon-terminated face is exposed to the growing SiO_2 . The fabrication of high-quality thermal oxides with low interface state and oxide-trap densities has proven to be a great challenge. Finally, the reliability of oxides is a major issue for SiC devices since, at high electric fields and high temperatures, oxides have poor longevity. This issue needs further research to reduce the leakage current in the devices that operate at elevated temperatures.

An important issue in high-temperature electronics is the type of metallization used, where examples include ohmic, Schottky, heat-sinking and capping.

It is necessary to have reasonable thermal expansion matching and good adhesion between the metal and SiC. The wide band gap of silicon carbide makes it difficult to control the electrical properties at the metal–semiconductor interface of devices. In addition, stable noncorrosive contacts are also key issues in high-temperature electronics. The main parameter of concern for SiC high-frequency devices is a stable Schottky barrier for good rectification and a low reverse leakage current while operating at elevated temperatures. Several groups have tried different combinations of transition metals that form good Schottky contact on n- and p-type SiC with barrier heights in the range $0.9\text{--}1.7 \text{ eV}$ [24.20, 39–41]. The rectifying properties either change to ohmic or degrade severely while operating at temperatures above 600°C . Among the ohmic contact the most widely used material for n-type is Ni_2Si which is generally formed by deposition of Ni film and silicidation is obtained by annealing at above 900°C . The Ni_2Si ohmic contact has been shown to be stable at very high temperatures [24.39, 42]. The formation of low-resistance ohmic contacts to p-type SiC is still difficult since metals with sufficiently large work functions are not available to offset the wide band gap and electron affinity of SiC. Aluminum is typically used to form p-type ohmic contacts. A major drawback of Al however is its relatively low melting point, which prohibits its use for high-temperature applications. Several other combinations of different metals have also been reported in the literature these have poor contact resistivities compared to Al [24.43, 44]. A special effort is required to develop stable contacts for SiC devices operating at higher temperatures, and metals with a high melting temperature and their silicides and carbides should be studied in the future towards this goal.

The packaging of SiC devices for high-power and high-frequency applications and operation at elevated temperature is an issue that has been neglected compared to material growth and device processing technology. It is highly desirable to find suitable packaging for high-temperature electronics which can endure high thermal stress and high power without the extra effort of cooling.

24.1.2 Diamond

Among the wide-band-gap semiconductors, diamond has the best properties, unmatched by any other material [24.45–49]. Most electrical, thermal and optical properties of diamond are extrinsic, i.e. strongly dependent on the impurity content [24.23, 46], the most common impurity being nitrogen. Diamond has

a large band gap (5.5 eV), high breakdown electric field (10 MV/cm), low dielectric constant (5.66–5.70), high carrier mobilities (≈ 1800 and $\approx 1600 \text{ cm}^2/\text{Vs}$ for electrons and holes, respectively [24.45]), high saturated carrier velocity ($2.7 \times 10^7 \text{ cm/s}$ and $1 \times 10^7 \text{ cm/s}$ for electrons and holes, respectively), high thermal conductivity (10–20 W/cmK), high resistivity (10^{13} – $10^{16} \Omega \text{ cm}$), low thermal expansion coefficient (1.1 ppm/K at room temperature), the highest sound velocity ($1.833 \times 10^6 \text{ cm/s}$), exceptional hardness ($10\,000 \text{ kg/mm}^2$) and wear resistance, low friction coefficient (0.05, dry), broad optical transparency [from 225 nm to far infrared (IR)], excellent resistance to radiation, chemical and thermal stability. A unique feature of diamond is that some of its surfaces can exhibit a very low or negative electron affinity. Obviously diamond is the material of choice for many applications, including electronics. The properties of diamonds make it the most suitable semiconductor for power electronics at high (RF) frequencies and high temperatures [24.50–53]. Since diamond-like silicon is a single-element semiconductor it is less susceptible to have the high density of structural defects that are usually present in compound semiconductors. However to date, diamond is regarded as one of the most difficult semiconductors to synthesize for the fabrication of electronic devices.

Diamond is a cubic semiconductor with lattice constant $a = 3.566 \text{ \AA}$. The covalent bonding of the carbon atoms (sp^3 bonds) is extremely strong and short, which gives diamond its unique physical, chemical and mechanical properties [24.46–49]. Diamond is available naturally and can also be synthesized. The natural form of diamond occurs as single crystals, whereas synthetic diamond can be prepared as single crystals, or as polycrystalline or nanocrystalline material. Usually natural diamond single crystals have a high nitrogen content and cannot be used for the fabrication of electronic components. The natural form of diamond has been classified according to several criteria, a detailed description of the classification of diamonds has been given by Walker [24.46] and Zajtsev [24.54].

Diamond melts at approximately 3827°C [24.45]. It is stable at elevated temperatures, but the stability depends on the ambient. In hydrogen ambient diamond is stable up to 2200°C [24.55], but it is graphitized in vacuum [24.56, 57] or in an inert gas [24.56]. Diamond does not have a native oxide, but it oxidizes in air at elevated temperatures. This is a critical point for the application of diamond for high-temperature devices. The oxidation of natural and synthetic diamond has been studied since the beginning of the 1960s; despite this

more research is needed for a complete understanding of the oxidation process. The activation energy for the oxidation of CVD-grown films in air was 213 kJ/mol for temperatures of 600 – 750°C and the oxidation proceeded by etching pits into the CVD film, thus creating a highly porous structure [24.58]. The results of several studies indicated that diamond oxidized preferentially. The oxidation of natural diamond and CVD-grown diamond films in oxygen has been observed to be dependent on the crystallographic orientation, here the (111) plane oxidized more easily than the (100) and (220) planes, and also the CVD films were less resistant to oxidation than natural diamond [24.56, 59]. Sun et al. [24.56] observed that the oxidation of synthetic diamond started in air at 477°C when oxygen is able to impinge into the densely packed (111) planes and they suggested that the oxidation of diamond occurs by the same mechanism as the corrosion of metals, whereby oxygen penetrates into the bulk by bonding and rebonding, leaving behind weakly interacting dipoles which are eroded away during processing. Lu et al. [24.60] reported that the oxidation in air of diamond films prepared by DC arc plasma jet started at 650°C which was about 100°C lower than the temperature of oxidation of natural diamond. Furthermore, it was reported that the oxidation rate of CVD diamond depended on the diamond's growth condition [24.61].

There are several etchants for diamond; the most commonly used method is oxidative etching. The effects of dry oxygen and a mixture of oxygen and water in the temperature range 700 – 900°C has been studied and compared with the effect of molten potassium nitrate [24.62, 63].

Diamond-based electronic devices have now been fabricated from natural and synthesized single crystals, high-purity single-crystal films (homoepitaxial diamond), and from polycrystalline films (heteroepitaxial diamond). Single crystals can be synthesized artificially by the high-pressure high-temperature method (HPHT), which mimics the process used by nature. The drawback of this method is that it produces single crystals limited in size. The largest crystals prepared by this method have a dimension of the order of millimeters, and the processing time to produce such crystal is very long [24.64]. These crystals have been used for the fabrication of discrete electronic devices and as substrates for homoepitaxial growth of diamond films by CVD technique. Diamond films have been epitaxially deposited on diamond single-crystal substrates, this demonstrates that single-crystal diamond deposition is possible by low-pressure processing [24.65, 66].

The discovery that diamond can be grown homoepitaxially and heteroepitaxially by the chemical vapor deposition (CVD) technique has opened up some of the expected applications of diamond. The history of this technique goes back to the late 1960s. During the 1980s, researchers [24.67, 68] made a series of discoveries which enabled them to grow, at significant growth rates, diamond films of high quality on non-diamond substrates by using hot-filament CVD and subsequently by microwave plasma chemical vapor deposition (MPCVD). This started a worldwide interest in diamond CVD for both research and technology. Since then a number of low-pressure CVD techniques have been developed [24.69, 70] and the volume of research on the preparation of large-area diamond films has been immense.

It has been shown recently [24.71, 72] that homoepitaxial growth of diamond films on high-quality HPHT diamonds can produce single-crystal films of a purity that exceeds the purity of the purest diamonds found in nature. The measurements of the carrier mobilities in these films have revealed interesting results, such as mobilities of 4500 and 3800 cm²/Vs for the electrons and holes, respectively. These values are the highest ever reported for diamond and are approximately twice as high as those found in pure natural diamond. The carrier mobility measurements were performed on homoepitaxial diamond deposited by a microwave plasma-assisted CVD technique and a HPHT diamond single crystal of dimensions 4 × 4 × 0.5 mm was used as the substrate. The homoepitaxial diamond film was found to be of exceptional purity and was found to contain a low concentration of intrinsic and extrinsic defects. The total measured nitrogen concentration was less than 1 × 10¹⁵ cm⁻³ and the dislocation density was less than 1 × 10⁶ cm⁻². The exceptionally high values of the carrier mobilities were attributed to the low defect and dislocation densities.

The disadvantage of diamond homoepitaxy is that only small-area single crystals can be fabricated, and substrates typically have a size of the order of millimeters. In order to exploit diamond's superior properties for the fabrication of electronic devices, thin diamond films are required, i.e. a method for the production of large-area, inexpensive single-crystal films with a low defect density.

Despite the progress made, the available diamond homoepitaxy methods cannot solve the technological problem of producing large-area diamond wafers for the fabrication of electronic devices. During the last 10 years film preparation has focused on diamond heteroepitaxy. The aim has been to produce films of homoepitax-

ial diamond's quality by avoiding the formation of grain boundaries and other defects. The research has focused on finding suitable substrates, conditions for achieving high diamond nucleation densities on various substrates, and the optimization of textured growth procedures. To date many substrates have been investigated, e.g. Ni, Co, Pt, Si, BeO, SiO₂, cubic BN, β -SiC, GaN, etc. [24.73–77]. Although some of these materials are suitable substrates, e.g. BN [24.77], all attempts to grow large diamond single-crystal films have hitherto failed. Most of the CVD diamond films reported to date have been grown on Si, mainly due to the availability of large-area single-crystal wafers and the low cost of Si as well as the favorable properties of Si [24.78–84]. These films are still polycrystalline but highly oriented (HOD) with respect to the substrate and have found application in many fields, e.g. electrochemical electrodes, field-emitter arrays, radiation detectors, micro-electromechanical systems (MEM), etc. In the field of MEMS a large number of devices for various applications have already been built, demonstrating thus the excellent properties of these films [24.85–87]. Despite their high quality, these films are not suitable for the fabrication of electronic devices since their attractive properties are deteriorated by structural imperfections, particularly by grain boundaries. The performance of electronic devices fabricated using such low-quality films is significantly reduced. Furthermore, due to their mosaic spread, these films cannot be used as substrates for the homoepitaxial growth of diamond.

A significant advance in diamond heteroepitaxy was made by the application of substrates with a multilayer structure. It was discovered that iridium single-crystal films grown as a buffer layer on MgO could serve as a substrate for the nucleation and growth of low-pressure microwave plasma-enhanced CVD diamond [24.88, 89]. The substrate MgO was later replaced by SrTiO₃ [24.90, 91], which decreased the mosaic spread of the epitaxial iridium and of the resulting heteroepitaxial diamond. The diamond layer had single-crystal quality and was used for the fabrication of field-effect transistors [24.92]. A further advance in the large-scale heteroepitaxial growth of diamond has been made recently when SrTiO₃ was replaced successfully by sapphire [24.93, 94]. Diamond produced in this way had the same high quality as that prepared on SrTiO₃. However, since sapphire is a relatively inexpensive large-area substrate, this development is a further step towards the wafer-scale production of heteroepitaxial diamond.

When cooling from the growth temperature the diamond film experiences significant compressive stresses

due to the difference in the thermal expansion coefficient of the materials present. These stresses can cause delamination, which is a serious obstacle for the development of diamond wafers. Moreover, calculations have shown [24.95] that films grown on MgO, SrTiO₃ and α -Al₂O₃ substrates are exposed to a significant amount of stress, -8.30 , -6.44 and -4.05 GPa, respectively. These large stresses make the treatment of thick films difficult. From the thermal stress point of view, Si with a stress of -0.68 GPa is the best substrate so far.

Further step towards the fabrication of large-area diamond single-crystal films for electronic applications has been made recently [24.95] by introducing a new concept for the substrate multilayer. The multilayer was prepared in two steps, where yttrium-stabilized zirconia (YSZ) was first deposited on Si, then an iridium thin film was deposited on the YSZ. This process decreases the lattice misfit between consecutive layers and such a substrate multilayer is shown in Fig. 24.5. The diamond was then grown on the iridium film as before. The quality of the diamond was the same as of that grown on SrTiO₃. The advantage of this concept is in the type and the combination of the substrate materials, which minimizes thermal stress thereby avoiding delamination.

In order to exploit diamond's superior properties for the fabrication of electronic devices, thin diamond films are required, i.e. a method for the production of large-area inexpensive single-crystal films with low

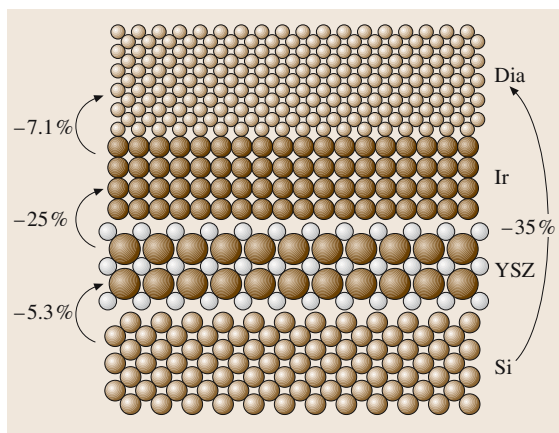


Fig. 24.5 Schematic representation of the layer system diamond/Ir/YSZ/Si(001). In the YSZ crystal the *large spheres* correspond to the oxygen ions. The *numbers* indicate the lattice mismatch between consecutive layers. (After [24.95]) (Figure provided by Mathias Schreck, University of Augsburg)

defect and dislocation densities is needed. Diamond device technology has similar problems to third-generation semiconductor technology, namely, the availability of inexpensive large-area diamond crystal with a low defect concentration; this is the prerequisite for a successful application of diamond for the fabrication of electronic devices.

24.2 Electronic Devices

24.2.1 Silicon Carbide

The ability of silicon carbide to operate at high temperature, high power, and high frequencies enables considerable enhancement of the performance of devices used within a wide variety of applications. In particular, SiC power devices can outperform equivalent Si devices, but this would require a mature and viable SiC semiconductor technology. The fundamental physical limitations of Si operation are the strongest motivation for switching to a wide-band-gap semiconductor such as SiC for high-temperature applications. The replacement of Si by SiC for power switches [24.20] is extremely advantageous, since the avalanche breakdown voltage for SiC is about ten times higher than that of Si (see Fig. 24.6 and Fig. 24.7). Moreover, silicon-carbide-based power switches also have a faster response with a lower parasitic resistance, so that the physical size of SiC devices

will be much smaller than equivalent silicon devices. In addition, the faster switching speed enhances the efficiency of power system conversion, and allows the use of smaller transformers and capacitors, which reduce significantly the overall size and weight of the system. The cooling requirements in power electronics, which are a considerable portion of the total size and cost of power conversion and distribution systems, can be significantly reduced by the high-temperature capabilities of SiC.

Power switching devices based on SiC such as Schottky barrier diodes, PiN junction diodes, metal-oxide-semiconductor field-effect transistors (MOSFETs), junction field-effect transistors (JFETs), bipolar junction transistors (BJTs) have already been demonstrated in the research stage. To date, SiC Schottky barrier diodes used for high-voltage applications have evolved from the research stage to limited commercial production [24.96]. Although Schottky barrier diodes offer rectification with

lower switching losses compared to PiN diodes (when switching from the conducting state to the blocking state) Si-based Schottky barrier diodes are still not used for high-voltage applications. The reason for this is the comparatively lower barrier height between ordinary metals and Si (typically less than 0.5 eV), which is further reduced in reverse-biased mode. The electron injection current from metal to semiconductor increases exponentially as the barrier height reduces. As a result very large reverse currents are observed at relatively low voltages, in the case of Si. The high Schottky barrier height of about 1.5 eV for SiC reduces the leakage current such that SiC Schottky diodes can operate at high voltages, and a blocking voltage of up to 5 kV has been demonstrated [24.39].

Diodes, Schottky and PiN Diodes

In contrast to Si diodes, SiC rectifiers exhibit ultimate low switching losses, no reverse current peak, and therefore an extreme soft recovery behavior. In Schottky diodes, a reduced forward-voltage drop and hence a reduced power loss would be desirable, and this can be achieved by lowering the Schottky barrier height. On the other hand, even at high temperatures, the barrier must be high enough to ensure a certain blocking voltage with a reasonably low reverse-leakage current. There is a strong dependency of the measured reverse current on the electric field for a Ti/SiC Schottky diode, with the temperature as a parameter [24.98].

Schottky barrier diodes with Ta and TaC on p- and n-type SiC have been investigated by a group at Carnegie

Mellon University [24.99]. The rectifying behavior of both Ta and TaC was observed on p-type 6H-SiC to 250 °C, while on n-type, TaC showed ohmic behavior above 200 °C [24.99]. The maximum operating temperature of SiC Schottky diodes is restricted because of the increasing leakage current and therefore, junction devices can approach and perform better at higher temperature.

The pn-junction diode characteristics at temperatures up to 400 °C have been reported both for epitaxially grown and ion-implanted SiC. The characteristics of the 6H-SiC pn-diode were obtained where nitrogen was implanted into p-type substrate. The rectifying ratio was measured to be 10^9 at room temperature and 10^5 at 400 °C. PiN diodes of 4H-SiC have been successfully fabricated and this device has a blocking voltage of up to 19 kV [24.100]. The high-temperature operation of a 8-kV diode indicates that the reverse leakage current density at 5 kV increases by only an order of magnitude between room temperature and 300 °C [24.101]. The reverse recovery of these diodes showed an extremely fast switching, which only increased by a factor of two between room temperature and 275 °C. This is small compared to Si diodes, which have an increase of four times between room temperature and 120 °C [24.101]. In the 5.5-kV blocking range of a PiN diode, a forward voltage drop of less than 5 V was observed at 500 A cm^{-2} between room temperature and 225 °C, in addition to a 50% increase in the peak reverse recovery current [24.102]. The reverse leakage current in a large-area ($6 \times 6 \text{ mm}^2$) diode which blocked 7.4 kV showed a small increase in the current up to 200 °C after packaging when measured up to 4.5 kV [24.103].

Another diode, the so-called junction barrier Schottky (JBS), is also considered an attractive device for power switching applications. The JBS device was first demonstrated in Si [24.104, 105]. The device structure is a combination of a Schottky barrier and a pn-junction, which allows a reduction in the power loss of the pn-junction under forward conduction and the utilization of the Schottky barrier. In the reverse direction the Schottky region is pinched off by the pn-junction, thus exhibiting a smaller leakage current. The spacing between the p^+ regions should be designed so that pinch-off is reached before the electric field at the Schottky contact increases to the point where excessive leakage currents occur due to tunneling, this complicates the device design and several attempts have been reported [24.106–108]. The major problem concerning SiC may be the poor ohmic contact with p-doped samples, which requires very-high-temperature

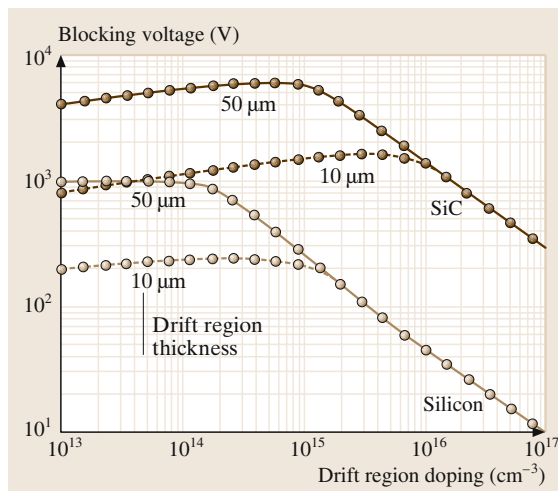


Fig. 24.6 Comparison of silicon carbide and silicon dielectric strength. (After [24.97])

annealing (above 850 °C) that causes severe damage to the Schottky contact; this then increases the excessive leakage current compared to an undamaged Schottky contact.

Though all these results on PiN diodes are very encouraging, there still remain problems in their practical operation. The most important perhaps, is the degradation of the current with time when operating under forward bias. One of the most attractive applications of SiC PiN diodes is in high-voltage DC (HVDC) electrical transmission systems. The current degradation is presumed to be related to the generation and extension of stacking fault defects in the basal plane. These defects lie in the crystalline plane perpendicular to the current flow direction. Recently it has been claimed [24,109] that this problem may be solved by growing the material along another direction. In order to get a defect-free crystalline structure, the material was grown on several *a*-face surfaces in at least 2–3 steps. The PiN diodes fabricated from this material were stressed and measured for 4 h at constant voltage. Growth using a few steps might be not feasible and more research on the operation of these PiN diodes is needed.

High-frequency MESFETs, SITs and BJTs

Concerning high-frequency devices, several groups from the USA, Japan and Europe have demonstrated two types of SiC transistors, metal–semiconductor field-effect transistor (**MESFETs**) and static induction transistors (**SITs**). These transistors suffer from large gate leakages, so that the possible application may be restricted to less than 350 °C. Static induction transistors (**SITs**) shows the best performance for high-peak-power applications up to 4 GHz. Under class C operation, the transistor produces over 350 W output power with 50% efficiency and a gain greater than 10 dB [24.110]. At ultra-high frequency (UHF) 900 W, pulsed power was obtained from a single chip packaged with a 51-cm gate periphery at L-band, and the same power was obtained for a 54-cm periphery package [24.110]. No investigations of the thermal limit of the high-temperature operation have been reported for such high-power **SITs**. For **MESFETs**, power densities as high as 5.2 W/mm² with 63% power added efficiency (**PAE**) have been demonstrated at 3.5 GHz while single device with 48 mm of gate periphery yielded 80 W continuous wave (CW) at 3.1 GHz with 38% **PAE** [24.111, 112]. Though these results are impressive, the transistor performance suffers severely from self-heating due to current flow in the device. Thus the majority of published articles are related to the small-gate periphery since thermal prob-

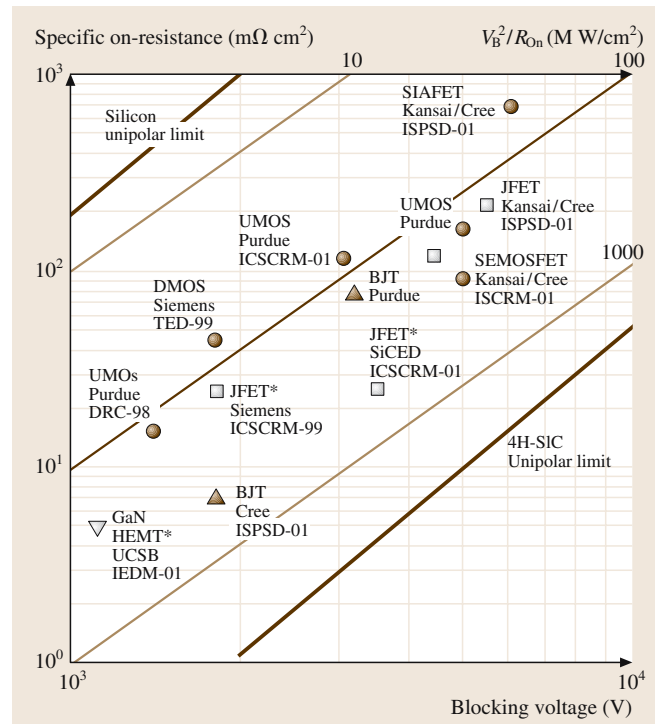


Fig. 24.7 Comparison of silicon and silicon carbide operating voltage and conduction resistance. (After [24.20])

lems are limited and the power densities are recorded to be very high. When the gate periphery increases, the power densities decrease very quickly, this is related to the self-heating effect. For a wide transistor, the temperature becomes very high and, therefore, the electron mobility decreases together with the drain current, and finally the **RF** power decreases. In a recent investigation, a channel temperature as high as 340 °C was observed for a large 31.5-mm transistor when only **DC** biases without any **RF** excitation was applied; the estimated dissipated power was 58.6 W in class AB [24.113]. In another investigation, the saturated **DC** drain current was reduced from 80 mA to 50 mA when the temperature was increased to 250 °C for a 6H-SiC **MESFET** [24.114].

Very little work has been published so far on bipolar transistors and specially within microwave applications. Due to the wide band gap, junctions of the transistors can control the reverse and forward current, and hence can be very attractive for high-ambient-temperature operations. However, the transistor suffers from some severe materials properties such as low minority-carrier lifetime, a very small hole mobility that restricts the cutoff (f_T), and a maximum frequency of oscillation (f_{max}).

The transistors have been fabricated and exhibited typical emitter breakdown at 500 V while the f_T was only 1.5 GHz. A transistor with an emitter width of 2.5 μm and an emitter periphery of 2.62 cm has demonstrated an output power of 50 W using a 80 V power supply in common-emitter class AB mode. The pulse width was 100 μs and the duty cycle was only 10%. The collector efficiency was 51% with a power gain of 9.3 dB [24.115]. Since SiC junctions can withstand high junction temperatures, the transistor can function efficiently without any external cooling, resulting in significant system advantages.

JFETs

SiC-based JFETs are very attractive for high-temperature electronics for the reasons described for BJTs. JFETs are more attractive as they are unipolar devices and thus do not suffer from a low value of the hole mobility. However, JFETs are normally depletion-mode (normally on) devices, and the gate must be kept at a negative voltage to keep the transistor off. Most power control systems require enhancement-mode (normally off) transistors so that the system can be switched off in a safe condition. One way to circumvent this problem is to connect a JFET in a cascode configuration with an enhancement-mode control device such as a Si or SiC MOSFET [24.116]. High-temperature operation of SiC-based JFETs has been reported in the literature since the early 1990s along with the evolution of SiC as a semiconductor for microelectronics.

Transistors with high blocking voltage up to 5.5 kV have been successfully fabricated. This transistor showed a specific on-resistance of 69 $\text{m}\Omega\text{ cm}^2$ and the turn off time was 47 ns [24.117]. The material 4H-SiC is more attractive due to high carrier mobilities which results in some favorable properties. Specifically, in a vertical JFET with the drain on the back side of the wafer, the forward current density reached 249 A/cm^2 at a drain voltage of only 1.2 V at room temperature, while at 600 °C the current dropped to 61 A/cm^2 with an increase of the specific on-resistance from 4.8 to 19.6 $\text{m}\Omega\text{ cm}^2$. In addition, the breakdown voltage of the transistor was 1644 V at room temperature, while it increased to 1928 V at 600 K [24.118]. In a thermal stress study of a 6H-SiC JFET, a decrease of about 40% in the drain current was observed at 300 °C [24.119].

High performance of SiC JFET has been reported by Siemens. The buried gate is permanently connected to the source, and the device blocked 1800 V at a specific on-resistance of 24 $\text{m}\Omega\text{ cm}^2$. Recently, Kansai Electric

and Cree Inc. reported the first enhancement-mode SiC JFET [24.120]. The device structure consisted of two gates; the buried gate was connected to the top gate, providing a gating effect from both sides. The transistor with a 50- μm -thick drift layer blocked 4.4 kV and the specific on-resistance was 121 $\text{m}\Omega\text{ cm}^2$.

MOSFETs

In contrast to Si MOSFETs, the 6H-SiC MOSFETs transconductance and channel mobility increases with rising temperature up to 225 °C. The high interface state density [24.121] may be the reason for these trends. MOSFETs fabricated from 6H-SiC have operated at temperatures around 400 °C with a very low drain leakage current in air [24.122]. In a recent report, a large-area device (3.3 mm \times 3.3 mm) blocked 1.6 kV with an on-resistance of 27 $\text{m}\Omega\text{ cm}^2$. This device exhibited a peak channel mobility of 22 cm^2/Vs and a threshold voltage of 8.8 V, which decreased to 5.5 V at around 200 °C [24.123].

The transistors were fabricated both in the lateral double-diffused MOSFET (DMOSFET) and the vertical U-shaped-trench MOSFET (UMOSFET) directions, these blocked several kilovolts. The Northrop Grumman group determined that both transistor types have similar channel mobilities and the same temperature effect due to the interface traps. In addition, they operated at up to 300 °C where the current and transconductance increased with rising temperature [24.124]. A group at Purdue University demonstrated a novel DMOSFET which could block 2.6 kV, where the drain current at 155 °C was four times the value at room temperature [24.125]. Recently, Purdue and Auburn University fabricated a UMOSFET that could block more than 5 kV, with a specific on-resistance of only 105 $\text{m}\Omega\text{ cm}^2$, for a 100- μm -thick low-doped drift layer (see e.g. Fig. 24.7). This device has not been characterized at elevated temperatures [24.126]. The transistor maintained a low on-resistance to current densities above 100 A/cm^2 , while the maximum reported current density was 40 A/cm^2 . Cree Research has already reported a record blocking of more than 6 kV.

Due to the problems caused by a high interface-state density, which results in a very low channel mobility, a new type of transistor, the accumulation-mode MOSFET (ACCUFET), with output characteristics that are similar to MOSFETs has been introduced and fabricated in both 6H- and 4H-SiC [24.127]. The 4H-SiC transistor exhibited a high specific on-resistance, 3.2 $\Omega\text{ cm}^2$ at a gate bias of 5 V, which reduced to 128 $\text{m}\Omega\text{ cm}^2$ at 450 K. The transistors were not de-

signed for high blocking voltages thus the unterminated breakdown voltage was only 450 V.

Recently, encouraging results from the continuing research into 4H-SiC MOSFET has been reported [24.128, 129]. Growing the gate oxide in N₂O ambient results in a significant enhancement of the inversion channel mobility in lateral n-channel Si face 4H-SiC MOSFETs. A mobility of 150 cm²/V s was obtained, whereas a value below 10 cm²/V s was obtained when growing the gate by the conventional process: in a wet or dry oxygen ambient.

Thyristors, GTOs and IGBTs

One of the most significant developments in device technology during the last 15–20 years is the insulated gate bipolar transistor (IGBT), which blocks high voltages but at the same time has a high conduction current. The basic advantage of the IGBTs is conductivity modulation due to carrier injection and a MOS-driven gate. The device operates between 600 and 6.5 kV where the current varies over 1–3500 A. The device has a big potential for further development in the areas of higher currents and voltages, and for higher frequencies and lower dissipated power.

A large amount of effort is being made on the development of SiC thyristors and IGBTs. The main advantage is the high electric breakdown field strength, which leads to very thin drift layers, and consequently much faster switching behavior.

The thyristor is the most popular controllable device used in high-power systems with controllable turn-on, and large-area Si thyristors are produced today as a single device on a wafer with a 4-inch diameter. The current capability is more than 1000 A with a blocking voltage approaching 10 000 V. These devices are used in high-voltage DC (HVDC) transmission systems, and so far no other device can match its performance. A 6H-SiC polytype thyristor was first demonstrated in early 1993. Due to the high resistivity of the p-type substrate, a high specific on-resistance of 128 mΩ cm² has been obtained for a 100 forward blocking voltage device [24.130]. However, at a higher temperature of 633 K, the thyristor performance improved with a specific on-resistance reduced to only 11 mΩ cm². The opposite polarity thyristor has been realized on the same structure with an improvement in all the characteristic parameters, and a very low specific on-resistance of about 3.6 mΩ cm² that increased to 10 mΩ cm² at 623 K temperature. Furthermore, early 4H devices exhibited a blocking voltage of –375 V. The low on-resistance of the 4H devices resulted in a lower voltage drop. Subse-

quently, the Northrop Gramman group has demonstrated a gate turn-off (GTO) with 1-kV blocking, but the device was turned-off by an external MOSFET. The device has been operated successfully at 390 °C [24.131]. Recently remarkable effort has been put into this area and 4H-SiC-based GTOs have been demonstrated with a 3.1-kV forward blocking capability with a 12-A conduction current [24.132].

24.2.2 Diamond

The many exceptional properties of diamond make it a very attractive material in several fields of electronics, optoelectronics and micro-electromechanical devices. Since an exhaustive review on the application of diamond in electronics and optoelectronics would be beyond the scope of this relatively short review, here the survey is limited to the application of diamond for pn-junctions, BJTs, FETs, passive components, heat-spreading elements, field emission, and resistors. For the application of diamond in other areas, e.g. sensors and detectors [24.133–139], microwave filters [24.140], acoustic wave filters [24.69, 141] and electro-mechanical microdevices [24.86, 142], the interested reader is directed to the above references.

pn-Junction

In order to exploit the superior properties of diamond for high-power/high-temperature electronic devices, high-quality inexpensive diamond must be available. The preparation of heteroepitaxial diamond is viable by bias-enhanced nucleation of iridium single-crystal film on sapphire or SrTiO₃ followed by low-pressure plasma-enhanced chemical vapor deposition. The crucial factor in the fabrication of electronic devices is the ability to prepare pn-junction, i.e. the ability to prepare in a reliable way p- and n-type materials with high carrier densities.

The p-type diamond has up to now been created by doping with boron. The boron is introduced from a gas or a solid [24.143–147] during diamond growth. The activation energy of electrical resistivity decreases with increasing boron concentration, being in the range 0–0.43 eV [24.143]. The energy is 0.35 eV for a boron concentration of 1×10^{18} cm⁻³ and it becomes zero for boron concentrations greater than 1.7×10^{20} cm⁻³ [24.143]. The p-type diamond can also be created by exposing the material to hydrogen plasma. The hydrogenation creates a layer with p-type conductivity with a fairly low resistivity $\approx 10^6$ Ω cm [24.148, 149]. The discovery that both single-crystal and poly-

crystalline diamond can be treated by the plasma has led to successful use of these materials for the fabrication of electronic devices, e.g. the fabrication of FETs.

The growth of n-type diamond is one of the most challenging issues in the diamond field. Many theoretical predictions as well as experimental attempts to obtain n-type diamond have recently been published. Unfortunately, all experimental attempts have up to now delivered n-type diamond with low electron densities; this restricts the use of n-type diamond for the fabrication of transistor. Nitrogen is often used to create n-type diamond, but nitrogen-doped diamond is an electrical insulator at room temperature due to the deep level of 1.7 eV of the nitrogen impurity. Nitrogen is thus not a practical dopant atom to use. Phosphorus would be an obvious candidate but it has an impurity level of ≈ 0.59 eV, which is still rather high. Several reports on phosphorus doping have been published reporting activation energies in the range 0.32–0.59 eV [24.143, 150–152]. Reproducible results on phosphorus doping have been reported recently [24.150, 151]. At room temperature a reasonably high electron mobility of $250 \text{ cm}^2/\text{V}$ was measured in the p-doped films [24.150]. Attempts to use other elements for n-type doping have also been made, e.g. sulphur (0.32 eV [24.143, 152, 153]), and lithium (0.16 eV [24.143, 154]). A survey of p-type doping of diamond has been given by Kalish [24.155].

Despite the experimental difficulties in n-type doping several research groups have succeeded to create a well function pn-junction. Koizumi et al. [24.150] created a pn-junction by epitaxial growth using a MPCVD technique to fabricate a phosphorus doped n-type diamond film on a boron doped p-type diamond film, on the [24.112] oriented surface of a diamond single crystal. The substrate contained over 100 ppm boron and exhibited high electrical conductivity. The electrical and optical properties of the junction have been investigated. The pn-junction exhibited clear diode characteristics, where the rectification ratio was over five orders in magnitude and the turn-on voltage was 4–5 V. The temperature dependence of the hole and electron concentrations in doped diamond films has been studied up to 1000 K. At room temperature the carrier mobilities of the boron and the phosphorus doped film were 300 and $60 \text{ cm}^2/\text{V s}$, respectively, and ultraviolet (UV) light emission has been observed at 235 nm.

A pn-junction has also been formed by using boron-doped p-type and nitrogen-doped n-type diamond films [24.156]. The active diamond films were grown by an MPCVD technique on heavily boron-doped HTHP

diamond single crystals. The forward and reverse I – V characteristics of the junction were studied as a function of temperature up to 400°C . At room temperature the resistivity of the n-layer is extremely high, $\approx 10 \text{ G}\Omega \text{ cm}$. The activation energy of the pn-junction diode saturation current obtained from these measurements was 3.8 eV. This value is in good agreement with theoretical predictions that assumes that the energy of the nitrogen level 1.7 eV. A pn-diode was also formed from a thick lithium-doped layer grown on a highly boron-doped substrate [24.143]. At room temperature the rectification ratio was $\approx 10^{10}$ for $\pm 10 \text{ V}$, but the series resistance was very high $\approx 200 \text{ k}\Omega$, whereas at higher temperatures the series resistance decreased to $30 \text{ k}\Omega$ but the rectification ratio decreased to $\approx 10^6$.

A single-crystal diamond pnp-type BJT consisting of boron/nitrogen pn-junctions has been fabricated [24.147, 156]. The transistor characteristics for common-base and common-emitter configurations have been measured. The temperature-activated leakage current limited the operation of the transistor up to 200°C . The study showed that it was possible to fabricate a pnp-type BJT, but the high resistance of the base $\approx 10 \text{ G}\Omega \text{ m}$ at 20°C limited the operation of the BJTs to the DC mode, small currents (nA), and moderate temperatures.

Diodes as electronic devices must have a high reverse voltage and a low resistance in the forward direction. It has been shown that lightly boron-doped (10^{17} cm^{-3}) Schottky diodes on oxygen-terminated surfaces exhibit good performance at high temperatures. Stable contacts at high temperatures have been developed to create diodes that may operate up to 1000°C [24.157–159]. The diodes showed breakdown behavior at reverse bias, but the reverse current was higher than theoretical values. The breakdown electric field for CVD diamond has been estimated to be 2.106 V/cm [24.143], a value well below that of natural diamond.

FETs

While the development of the diamond bipolar junction transistor is hindered by the lack of n-type diamond with high electron densities, there has been some promising results towards the development of diamond field-effect transistors, although studies at high temperatures are absent. Nevertheless, a short review of the present situation regarding diamond FETs is given here. There are presently two concepts for obtaining high-performance devices: boron δ -doped p-channel FETs and hydrogen-induced p-type surface-channel FET. Since a shallow diamond dopant is absent, δ -doping and

two-dimensional (2D) conduction are essential parts of the diamond device concept.

Hydrogenated surfaces of natural diamond and films grown by plasma-assisted chemical vapor deposition technique exhibit substantial conductivity. Hydrogenation is carried out by treating the diamond surface with hydrogen plasma and cooling to room temperature in hydrogen ambient. The hydrogen-terminated surfaces exhibits the following, a p-type conductivity (2D conduction) of the order of 10^{-4} – $10^{-5} \Omega^{-1}$, a shallow acceptor level which is less than 50 meV, a high carrier concentration 10^{13} cm^{-2} , and a hole mobility in the range 100–150 cm^2/Vs [24.148, 149, 161–165]. The thickness of the conductive layer is estimated to be less than 10 nm. It has been suggested that the surface conductivity depends on the details of the hydrogen termination together with the coverage of physisorbed adsorbates [24.166–168]. The high conductivity of the layer can be destroyed by dehydrogenation of the surface, by exposing the surface to oxygen or by heating in air at a temperature above 200 °C [24.148, 169]. The latter consideration may complicate the fabrication of devices. In order to exploit the conducting layer for the fabrication of high-temperature diamond devices, a way of protecting the layer has to be developed [24.169]. The opposite of hydrogenation is obtained by treating the diamond surface with oxygen plasma. The oxygen-terminated surface becomes highly insulating due to the surface potential pinning at $\approx 1.7 \text{ eV}$ above the valence band [24.170]. This effect is exploited to isolate parts of a device when fabricating devices using the planar process. The hydrogen-terminated areas are transformed into oxygen-terminated ones. The hydrogen-terminated surface has a low surface-state density thus the barrier height between a metal and the surface is almost determined by the work function of the metal only. Usually Al and Au are used as ohmic contacts. Furthermore, when hydrogenation is used to create surface p-type conductivity, alternative cheaper types of diamond can be used [24.171].

The high hole conductivity allows for the fabrication of high-performance FETs, e.g. [24.160, 162, 172–175]. Two types of FET have been investigated, the metal–insulator–semiconductor field-effect transistor (MISFET) and the metal–semiconductor field-effect transistor (MESFET). The high-frequency performance and DC output characteristics of a MISFET with a 0.7- μm gate length has been investigated at room temperature [24.175]. The measurements show that the cutoff frequency f_T and the maximum frequency (MAG – maximum available gain) f_{max} are 11 and

18 GHz, respectively. In addition, a transconductance, g_m , of 100 mS/mm, has been obtained from DC measurements. The DC output characteristics of a circular FET with a gate length of 3 μm has been investigated at room temperature [24.176]. The maximum current was 90 mA/mm at a gate bias -6 V , and the maximum transconductance was 25 mS/mm. Breakdown voltage at pinch-off was -200 V . The high-frequency performance of MESFETs has also been also investigated at room temperature [24.177]. The cutoff frequency and the maximum frequency for a transistor with a 2- μm gate length were 2.2 and 7.0 GHz, respectively. The highest frequencies at room temperature were obtained for a MESFET with a gate length of 0.2 μm [24.174, 178, 179], and the following values were reported: maximum drain current, I_{Dmax} , 360 mA/mm, the transconductance, g_m , 150 mS/mm, the maximum drain voltage, V_{DSmax} , 68 V, and the maximum estimated power capability, P_{Rfmax} , 3.0 W/mm for class A.

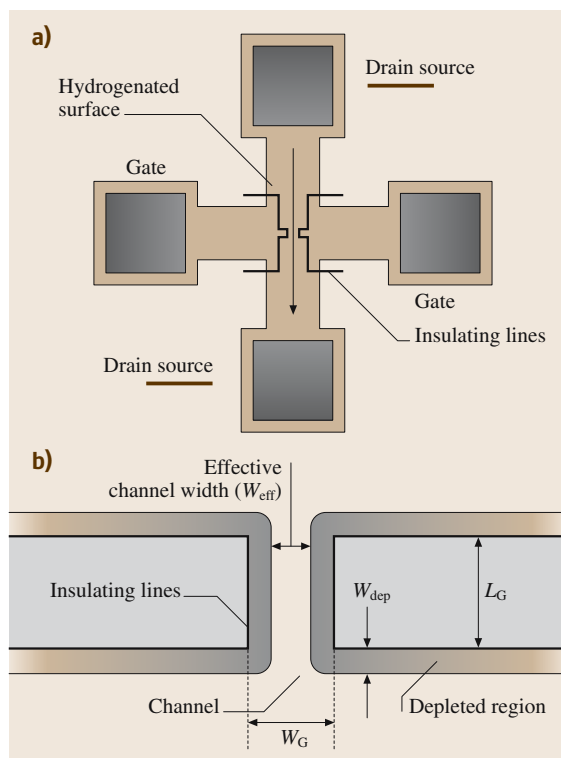


Fig. 24.8 (a) Schema of an in-plane transistor fabricated on a H-terminated diamond surface, using oxidized lines to define channel and gate, (b) channel region with oxidized lines and depletion regions along the lines. L_G and W_G are the gate length and gate width, respectively. (After [24.160])

The cutoff and maximum frequencies were measured to be 11.7 GHz and 31.7 GHz, respectively [24.178]. In a later investigation even higher frequencies were obtained, 21 GHz and 63 GHz, respectively [24.179]. These results indicate that a high gain can be obtained and that this transistor can operate at RF frequencies. Unfortunately, the instabilities of the hydrogen-terminated surface prevented large signal measurements on the device. The structure of a FET fabricated using the hydrogen layer is shown in Fig. 24.8.

The hole concentration of homoepitaxially prepared diamond films was measured for temperatures up to 500 °C. By altering the manner in which the temperature is increased and decreased, it has been shown

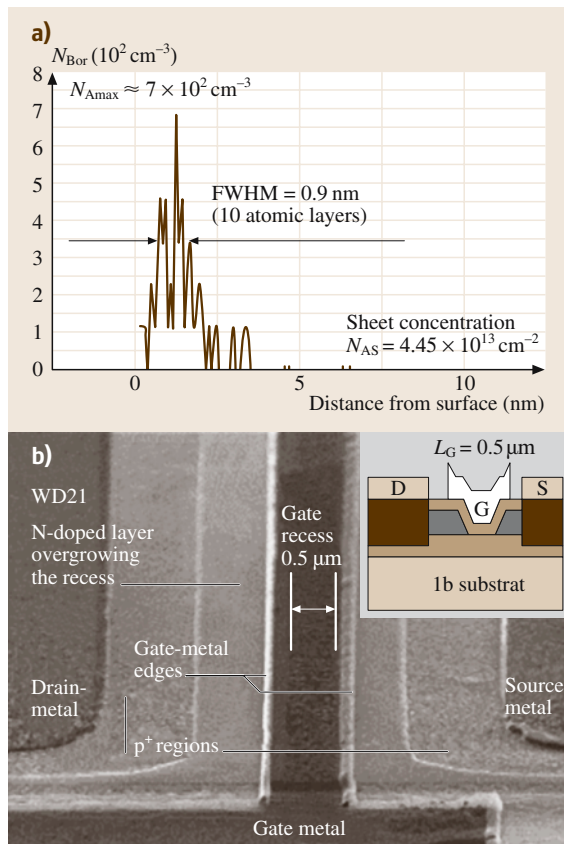


Fig. 24.9 (a) Elastic recoil detection (ERD) analysis of a δ -doped layer showing a full-width at half-maximum (FWHM) of 0.9 nm (10 atomic layers); (b) fabricated recessed gate δ -channel JFET with roughness visible at the bottom of the epitaxially overgrown recess trench. The gate length of 0.5 μm is given by the width of the gate recess. (After [24.174])

that adsorbates from the environment can change the hole concentration in the conducting layer. Furthermore, the influence of the crystalline and surface properties of diamond homoepitaxial layers on device properties of hydrogen-terminated surface-channel FETs has been investigated [24.180].

The δ -channel FET approach is based on boron δ -doping, where the doping profile with peak concentration of 10^{20} cm^{-3} must be confined to within a few nanometers [24.181,182]. The DC output characteristics of δ -doped MESFETs were investigated at room temperature and at 350 °C [24.183]. The maximum drain current I_{Dmax} at room temperature and 350 °C were 35 $\mu\text{A}/\text{mm}$ and 5 mA/mm, respectively. The maximum usable drain-source voltage V_{DSmax} at room temperature and 350 °C were 100 V and 70 V, respectively. A δ -channel JFET, with a δ -doping profile width of 1 nm, has been fabricated and the DC output performance has been investigated at a temperature of 250 °C and −125 °C [24.174]. The mobility of such δ -doped channels would be $\approx 350 \text{ cm}^2/\text{V s}$. At temperature 250 °C all the carriers are activated and a maximum drain current I_{Dmax} of 120 mA/mm has been measured for a transistor with a gate length of 2 μm and a gate width of 100 μm . Figure 24.9 shows δ -doping profile and fabricated recessed gate δ -channel JFET.

All FETs were fabricated from homoepitaxial films deposited on HTHP single-crystal diamond substrates, since the use of polycrystalline diamond films significantly reduced the FETs' performance. In addition, encouraging results which can be regarded as an important breakthrough in the development of high-power diamond electronics on the wafer scale, have been reported recently. The fabrication of high-performance p-type surface-channel MESFETs with sub-micron gate length from a heteroepitaxially grown diamond film. The diamond film was grown on a SrTiO_3 ceramic substrate with an intermediate iridium buffer layer [24.92]. The components exhibited similar performance as structures fabricated from homoepitaxial diamond films. The devices have been analyzed at room temperature under DC, small signal and large-signal RF power conditions. For devices with the smallest gate length of 0.24 μm the following values were obtained, $I_{\text{Dmax}} \approx 250 \text{ mA}/\text{mm}$, transconductance $\approx 97 \text{ mS}/\text{mm}$ and $V_{\text{Dmax}} = 90 \text{ V}$. The cutoff f_T and maximum frequencies $f_{\text{max(MAG)}}$ were measured to be 9.6 and 16.3 GHz, respectively.

In connection with the solid-state diamond FETs, the diamond vacuum field-effect transistor (VFET) ought to be mentioned. This component is part of vacuum electronics using diamond electrodes [24.184]. The diamond

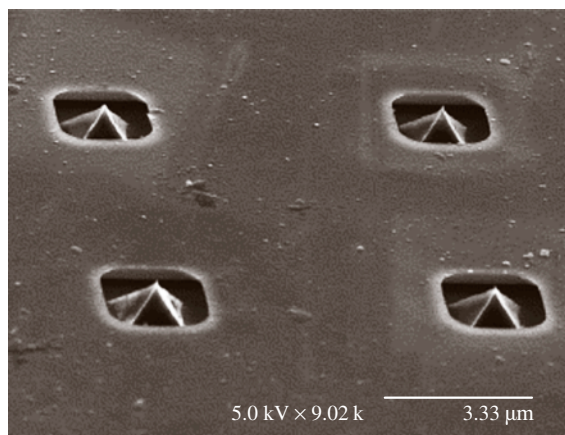


Fig. 24.10 Scanning electron microscopy (SEM) picture of a 2×2 array of gated diamond emitters. (After [24.185])

field emitter has been integrated with silicon-based micro-electromechanical system processing technology in order to obtain a monolithic diamond field-emitter triode on a Si wafer [24.185] (see Fig. 24.10). The monolithic VFET with a self-aligned n^{++} gate has been operated in a vacuum in the triode configuration, with an external anode placed $50\text{ }\mu\text{m}$ above the gate. The low turn-on gate voltage of 10 V, the high anode emission current of $4\text{ }\mu\text{A}$ at an applied gate voltage of 20 V, and the high gain factor of 250, indicate that high-speed diamond VFETs for low and high power can be realized.

Passive Components

Highly oriented heteroepitaxial diamond films grown on wafer-size Si, HOD films, are polycrystalline and cannot therefore be used for the fabrication of transistors or as substrates for the homoepitaxial growth of diamond. Nevertheless, these films possess properties that approach the properties of ideal diamond, and may be the most suitable material for MEMS technologies, heavy-duty, high-frequency and high-temperature electronic components for integrated high-power microwave circuits, and passive components like e.g. capacitors, switches, and planar waveguides. This should also enable the integration with Si- and 3C-SiC electronics.

In connection with high-temperature devices two passive components can be mentioned here, capacitors [24.187] and switches [24.188]. The capacitor has been realized using a diamond membrane as the dielectric and Au as the contacts. The diamond films were grown on large-area Si substrate by the MPCVD technique. The dielectric loss of the capacitor has been measured up to 600°C and the C - V characteristics for

up to 500°C . The capacitor functioned up to 450°C with low loss and constant capacitance, whereas for higher temperatures the capacitance was temperature-dependent. The micro-switch was fabricated using the same type of diamond. The properties of diamond make it an ideal material for switch fabrication. Switches often operate in a vacuum at temperatures around 650°C in a kHz frequency range. Moreover, no change in the switching threshold voltage indicates that there has been no change in the device's mechanical properties.

Heat-Spreading Elements

The properties of diamond suggest that its main application will be in power electronics at high frequencies and high temperature. Natural diamond has the highest known thermal conductivity, 2200 W/mK at room temperature. The highest-quality synthetic diamond grown by CVD technique has an identical thermal conductivity. This implies that power devices can be built on an integrated heat spreader. The high thermal conductivity and thermal stability of diamond will allow ideal heat sinks to be combined with high-power microelectronic and optoelectronic devices operating at high temperatures. Although the thermal conductivity of CVD diamond reduces with decreasing layer thickness, it still exceeds the conductivity of gold, even when the thickness is as low as $2\text{ }\mu\text{m}$ [24.189, 190].

GaN-based FETs have, owing to their excellent characteristics, great potential for high-frequency power

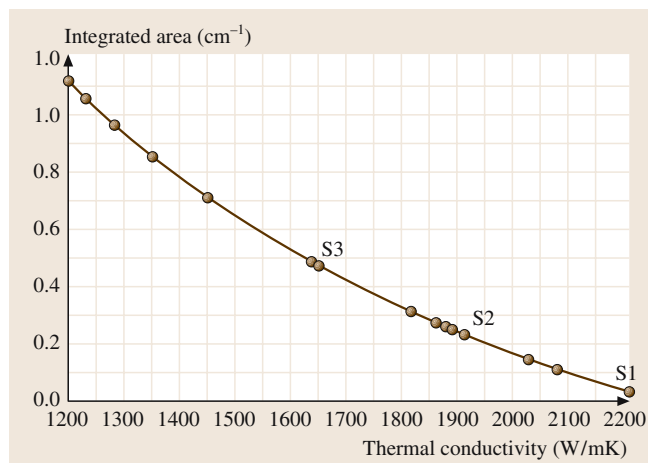


Fig. 24.11 The measured integrated absorption between 2760 and 3030 cm^{-1} plotted against the thermal conductivity determined using the laser flash technique. Samples S1–S3 are marked, the other data points are measured from a range of material with as-grown thickness of 100 – $800\text{ }\mu\text{m}$. (After [24.186])

amplifiers [24.21]. A limitation of the performance of GaN FETs is the rise in operating temperature caused by heat dissipation, which leads to large leakage currents and reduced channel mobility [24.191, 192]. It has been demonstrated that a heat-spreading diamond film can be deposited on GaN FETs, without any degradation of the transistors characteristics [24.190]. The films of thickness 0.7 and 2 μm were deposited by the CVD technique using a new seeding process at low temperatures (less than 500 °C). The GaN FETs were fabricated by a standard process for GaN transistors.

A correlation between the integrated absorption in the IR range of 2700–3030 cm^{-1} , associated with stretch modes of CH_4 , and the measured thermal conductivity at 300 K were reported for CVD diamond samples of different optical quality [24.186]. This provides a quick and reliable method of ascertaining a film's thermal quality. The measured integrated absorption as a function of the thermal conductivity is shown in Fig. 24.11.

Field Emission

The emission of electrons from an electrically active surface by means of tunneling into vacuum is commonly referred to as field emission. The emission properties of diamond, diamond-like carbon (DLC) and carbon-based materials are very robust [24.184]. Thin diamond film deposited by the CVD technique is an excellent cold electron emitter. The electron density yield is around 1 mA/cm^2 for an applied field of 10–100 $\text{V}/\mu\text{m}$. There is much scientific and technological interest in electron field emission from carbon-based materials for applications in many fields, e.g. vacuum field-effect transistors, diodes and triodes, ion sources, electron guns, flat-panel displays, scanning microscopes, energy conversion, and many others. A very large number of reports concerning the field emission from carbon-based materials has been published since the first reported electron emission from diamond in 1991 [24.193]. The physical reason behind the outstanding emission properties of these materials however is still under discussion [24.194].

Depending on the surface termination, diamond can exhibit a low or even negative electron affinity (NEA) [24.165]. The surfaces of diamond and DLC prepared by the CVD technique are naturally terminated by physisorbed hydrogen, making diamond the only semiconductor for which a true negative electron affinity can be obtained. This means that the vacuum level can be below the conduction-band minimum at the surface barrier [24.165]. The hydrogen layer lowers the electron affinity and reduces the turn-on voltage re-

quired to achieve field emission. Many reports treating different aspects of the field-emission process has been published to date, e.g. [24.195–197]. However, there are several problems connected with a direct application of diamond, DLC or other forms of carbon emitters, for instance nonuniform emitter microstructures that result in inconsistent emission and poor long-term stability. In diamond there are no free electrons to be easily emitted. A three-step emission model was proposed by Cutler et al. [24.198]. According to this model the electrons are injected through the back contact, transported through the diamond, and emitted at the diamond–vacuum surface. This model was extended by a proposition [24.199] that: (i) electrons are transferred by the applied electric field from the substrate to the nitrogenated DLC conduction band through a band-to-band tunneling process; (ii) electrons are transported across the nitrogenated DLC film through its conduction band, and (iii) electrons are emitted into the vacuum at the surface of nitrogenated DLC film by tunneling through the barrier. The barrier was crated by the strong upward band-bending at the surface.

Hydrogen has been the most commonly used element for inducing NEA on diamond surfaces. It has been reported that other materials e.g. a Ti layer, can also induce NEA on diamond surfaces [24.200]. High-temperature thermionic electron emission has been studied for several decades [24.201]. Fowler–Nordheim theory, the theoretical foundation of field emission, neglects the effects of temperature on emission characteristics. The temperature dependence of field emission from diamond film surfaces has been reported recently. The electron emission properties of nitrogen-doped diamond films were studied as a function of temperature for up to 950 °C [24.202]. The films were prepared by a microwave-assisted CVD technique for application as a low-temperature thermionic field-emission cathode. The film surfaces were terminated with hydrogen or titanium. Measurements at elevated temperatures showed the importance of a stable surface passivation. Hydrogen-passivated films showed enhanced electron emission, but measurements at elevated temperatures, e.g. 725 °C, showed that the hydrogen layer had degraded rapidly. The titanium-terminated films showed a similar enhanced emission as the hydrogen-terminated ones but the titanium-passivated surfaces were stable up to 950 °C. The electron emission increased with increasing temperature, implying that the field emission is strongly temperature-dependent. At temperatures below 500 °C no emission was detected, while increasing the temperature above 700 °C a strong contribution to

the electron emission was detected. At constant temperature the emission current was nearly constant at low anode voltages, whereas it increased exponentially for anode voltages above 15 kV. At low anode voltages, below 10 kV, the activation energy was approximately 1 eV. This energy value indicates contributions from nitrogen donor levels and defect states to the emission current. It was suggested, according to the emission characteristics, that the emission at low anode voltages could be attributed to thermionic emission of electrons in the diamond's conduction band, whereas the exponential increase at higher voltages indicated a tunneling through or thermionic emission over the potential barrier.

The temperature dependence of electron field-emission characteristics of nitrogen-doped polycrystalline diamond films has also been investigated by [24.203]. The films were grown by the PECVD technique for high-temperature applications. The measurements show that maximum current density increased and the turn-on voltage decreased for temperatures between 300 and 500 °C. The experimental results were interpreted in terms of the Fowler–Nordheim theory and by means of an alternative model that included size and temperature effects explicitly. Furthermore, analysis of the temperature dependence of emission was carried through by parameter estimation of the effective emitting area, field-enhancement factor, and work function. All of the estimates indicated that the emission characteristics exhibit a strong dependence on temperature. From these results it was suggested that the thermally excited electrons are responsible for improved emission at high temperature.

Resistors

An attractive feature of diamond films is their ability to be an excellent dielectric when undoped, or as an interesting resistors when doped. Even polycrystalline or nanocrystalline diamond films, depending on

the process deposition conditions, exhibit breakdown strength (as a dielectric) and power-density capability (as a resistor) with values particularly interesting for high-temperature and high-power applications. Diamond resistors were fabricated on a ceramic substrate of aluminum nitride [24.204], by etching in oxygen plasma from the CVD-deposited layer of boron-doped diamond films. The resistors were intended for investigation of high-power-density characteristics of boron-doped diamond. The resistor exhibits ohmic behavior at low to medium current levels, e.g. up to ≈ 3 mA. At higher currents, thermal excitations are important, and the carrier-density enhancement and conductivity increases. The component enters a thermal runaway situation, but unlike conventional devices, the resistor by virtue of diamonds tolerance for high power, continues to operate at very-high-power and temperature conditions. Operating the resistor under load at high voltages is not unlike operating certain devices in the reverse breakdown mode, except for the much higher power-density levels. For example, the resistance at low power is typically 315 k Ω . The power density just before entering the thermal runaway region is 480 kW/cm². At a maximum current of 14.2 mA and voltage of 413.2 V the resistance fell to 29 k Ω . At the end of the test the power was 5.87 W and the current density was 4730 A/cm² [24.204].

Technological applications of micro-electromechanical systems (MEMs) and optoelectronic system can be broadened by the existence of thermally stable thin-film resistive microheaters. Thermally stable resistors working in a wide temperature range were presented recently [24.205]. Metal containing amorphous carbon like diamond (DLC) films, CrSiDLC and MoSiDLC, were prepared by the CVD technique and doping with Mo and Cr. The resistors, in the form of strips, were deposited on glass ceramic substrates. The films were studied at 700–800 °C. The average resistivity was $2 \times 10^{-3} \Omega$. The thermal coefficient of resistance was $7 \times 10^{-4} \Omega/K$ and the activation energy was 30 meV.

24.3 Summary

This chapter reviewed the advances in research and technology of the wide-band-gap semiconducting material SiC and diamond for high-temperature electronics. The hexagonal 4H-SiC structure is the most appealing for high-temperature power electronics. Basic SiC devices for use in power electronics that can operate at elevated temperature were presented. Four major de-

vice concepts: Schottky and pn-diode, high-frequency devices, power switching devices and high-current devices were discussed. PiN high-blocking-voltage power diodes can operate above 300 °C with an order of magnitude increase in the reverse leakage-current density. JFET is supposed to be the most promising power device. Transistors with blocking voltage up to 5 kV were

characterized up to 300 °C. MOSFETs are still under development, and their operation has been tested up to 400 °C. The operation of high-frequency transistors are limited by to Schottky barrier height and possible applications may be limited to below 350 °C. The most probable near-future devices are thyristors and GTOs since their junctions can withstand higher temperatures, and exhibit large blocking voltages and high conduction currents.

Concerning high-temperature/high-power electronics, diamond has the best properties and is at different stage of research than SiC, since the technology of diamond electronic devices is in its infancy. The foremost progress in diamond research has been in the preparation of high-quality heteroepitaxial diamond, which is the prerequisite for wafer-scale fabrication.

The state of the art concerning high-temperature devices are pn-diodes and Schottky diodes that operate up to 400 °C and 1000 °C, respectively, and passive com-

ponent capacitors and switches that operate up to 450 and 650 °C, respectively. Moreover, the development of FETs have reached important milestones for both DC and RF modes. The performance of these components, however, are still below standard, and components that can operate at high temperatures and high powers are still far away. The development of the BJT is hindered by the unavailability of n-type diamond with high electron concentrations. The application of diamond for field emission has made a lot of progress, since vacuum FETs have been fabricated and field emission from diamond has been studied at high temperatures. Diamond as a heat sink for high-power electronic devices can soon find wide applications due to the recent progress in the preparation of large-area high-quality diamond films. Carbon-based resistors exhibit excellent thermal stability, and with the continuing progress in the preparation of carbon-based materials, will find increasing application in the fabrication of high-temperature devices.

References

- 24.1 P. G. Nuedeck, R. S. Okojie, L-Y. Chien: Proc. IEEE **90**, 1065–1076 (2002)
- 24.2 P. L. Dreike, D. M. Fleetwood, D. B. King, D. C. Sprauer, T. E. Zipperian: IEEE Trans. A **17**, 594–609 (1994)
- 24.3 C. S. White, R. M. Nelms, R. W. Johnson, R. R. Grzybowski: High temperature electronic systems using silicon semiconductors, Industry Applications Conference, 1998. Thirty-Third IAS Annual Meeting, Proc. IEEE **2**, 967–976 (1998)
- 24.4 M. Willander, H. L. Hartnagel (Eds.): *High Temperature Electronics* (Chapman Hall, London 1997)
- 24.5 S. Cristoloveanu, G. Reichert: High-Temp. Electron. Mater. Dev. Sensors Conf., 86–93 (1998)
- 24.6 B. Gentinne, J. P. Eggermont, D. Flandre, J. P. Colinge: Mater. Sci. Eng. B **46**, 1–7 (1997)
- 24.7 L. Demeus, P. Delatte, V. Dessard, S. Adriaensen, A. Viviani, C. Renaux, D. Flandre: The art of high temperature FD-SOI CMOS, IEEE Cat. No. 99EX372, IEEE Conference Proceedings (1999), High Temperature Electronics, 1999. HITEN 99. The Third European Conference on 4–7 July 1999, pp. 97–99
- 24.8 S. Cristoloveanu: Circ. Dev. Mag. **15**, 26–32 (1999)
- 24.9 J. C. Zolner: Solid-State Electron. **42**, 2153–2156 (1998)
- 24.10 P. Schmid, K. L. Lipka, J. Ibbetson, N. Nguyen, U. Mishra, L. Pond, C. Weitzel, E. Kohn: IEEE Electron. Dev. Lett. **19**, 225–227 (1998)
- 24.11 I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan: J. Appl. Phys. **89**, 5815–5875 (2001)
- 24.12 C. M. Carlin, J. K. Ray: The requirements for high-temperature electronics in a future high speed civil transport, Trans. 2nd Int. High-Temperature Electronics Conf., Charlotte, NC June 1994, ed. by D. B. King, F. V. Thome (Phillips Laboratory, Sandia National Laboratories, Wright Laboratory, 1994) 1.19–1.26
- 24.13 Z. D. Gastineau: High-temperature smart actuator development for aircraft turbine engines, Proc. 5th Int. High-Temperature electronics Conf., Albuquerque, NM, June 2000, pp. X.1.1–X1.5
- 24.14 S. P. Rountree, S. Berjaoui, A. Tamporello, B. Vincent, T. Wiley: High-temperature measure while drilling: Systems and applications, proc. 5th Int. High-Temperature Electronics Conf., Albuquerque, NM, June 2000, pp. X.2.1–1.2.7
- 24.15 M. Gerber, J. A. Ferreira, I. W. Hofsaier, N. Seliger: High temperature, high power density packaging for automotive applications, Power Electronics Specialist Conference, 15–19 June, 2003. PESC '03. 2003 IEEE 34th Annual IEEE Vol **1**, 425–430 (2003)
- 24.16 S. Lande: Supply and demand for high temperature electronics, High Temperature Electronics, 1999. HITEN 99. The Third European Conference on 4–7 July, 1999, pp. 133–135
- 24.17 J. B. Baliga: IEEE Electron. Dev. Lett. **10**, 455–457 (1989)
- 24.18 A. R. Powell, L. B. Rowland: Proc. IEEE **90**, 942–955 (2002)
- 24.19 A. Elasser, T. P. Chow: Proc. IEEE **90**, 969–986 (2002)
- 24.20 J. A. Cooper, A. Agarwal: Proc. IEEE **90**, 956–968 (2002)
- 24.21 R. J. Trew: Proc. IEEE **90**, 1032–1047 (2002)

- 24.22 J. A. Cooper: Mater. Sci. Forum **389–393**, 15–20 (2002)
- 24.23 J. F. H. Custers: Physika **18**, 489–496 (1952)
- 24.24 L. S. Pan, D. R. Kania (Eds.): *Diamond: Electronic Properties and Applications* (Kluwer Academic, Boston 1995)
- 24.25 J. Walker: Rep. Prog. Phys. **42**, 1605–1660 (1979)
- 24.26 <http://www.eetimes.com/story/OEG20030822S0005>
- 24.27 J. Palmour, R. Singh, R. C. Glass, O. Kordina, C. H. Carter: (1997), Silicon carbide for power devices, Power Semiconductor Devices and ICs 1997, ISPSD '97, 1997 IEEE International Symposium on, 26–29 May 1997, pp. 25–32
- 24.28 A. Ellison: PhD Thesis; Dissertation No. 599, Linköping Studies in Science and Technologies, Linköping University (1999)
- 24.29 G. R. Fisher, P. Barnes: Philos. Mag. **B61**, 217–236 (1990)
- 24.30 J. A. Lely: Ber. Dt. Keram. Ges. **32**, 299 (1955)
- 24.31 Y. M. Tsirov, V. F. Tsvetkov: J. Cryst. Growth **43**, 209–212 (1978)
- 24.32 D. Nakamura, I. Gunjishima, S. Yamagushi, T. Ito, A. Okamoto, H. Kondo, S. Onda, K. Takatori: Nature **430**, 1009–1012 (2004)
- 24.33 P. G. Nuedeck, J. A. Powell: IEEE Electron. Dev. Lett. **15**, 63–65 (1994)
- 24.34 H. McD. Hobgood, M. F. Brady, M. R. Calus, J. R. Jenny, R. T. Leonard, D. P. Malta, S. G. Müller, A. R. Powell, V. F. Tsvetkov, R. C. Glass, C. H. Carter: Mater. Sci. Forum **457–460**, 3–8 (2004)
- 24.35 A. R. Powell, R. T. Leonard, M. F. Brady, S. G. Müller, V. F. Tsvetkov, R. Trussel, J. J. Sumakeris, H. McD. Hobgood, A. A. Burk, R. C. Glass, C. H. Carter: Mater. Sci. Forum **457–460**, 41–46 (2004)
- 24.36 R. C. Glass, L. O. Kjellberg, V. F. Tsvetkov, J. E. Sundgren, E. Janzén: J. Cryst. Growth **132**, 504–512 (1993)
- 24.37 J. P. Bergman, H. Lendermann, P. A. Nilsson, U. Lindefelt, P. Skytt: Mater. Sci. Forum **353–356**, 299–302 (2001)
- 24.38 J. R. Jenny, D. P. Malta, M. R. Calus, S. G. Müller, A. R. Powell, V. F. Tsvetkov, H. McD. Hobgood, R. C. Glass, C. H. Carter: Mater. Sci. Forum **457–460**, 35–40 (2004)
- 24.39 H. M. McGlothlin, D. T. Morissette, J. A. Cooper, M. R. Melloch: 4 kV silicon carbide Schottky diodes for high-frequency switching applications, Dev. Res. Conf. Dig. 1999 57th Annual, 28–30 June 1999, Santa Barbara, California, USA, pp. 42–43
- 24.40 S. Liu, J. Scofield: Thermally stable ohmic contacts to 6H- and 4H- p-type SiC, High Temperature Electronics Conference, HITEC 1998, 4th International, 14–18 June 1998, pp. 88–92
- 24.41 Q. Wahab, A. Ellison, J. Zhang, U. Forsberg, E. A. Duranova, L. D. Madsen, E. Janzén: Mat. Sci. Forum **338–342**, 1171–1174 (1999)
- 24.42 R. Raghunathan, B. J. Baliga: IEEE Electron. Dev. Lett. **19**, 71–73 (1998)
- 24.43 L. Zheng, R. P. Joshi: J. Appl. Phys. **85**, 3701–3707 (1999)
- 24.44 K. J. Schoen, J. P. Henning, M. Woodall, J. A. Cooper, M. R. Melloch: IEEE Electron. Dev. Lett. **19**, 97–99 (1998)
- 24.45 Landolt-Börnstein: *Condensed Matter Group III/41A2, Impurities and Defects in Group IV Elements, IV–IV and III–V Compounds. Part α : Group IV Elements*, ed. by W. Martienssen (Springer, Berlin, Heidelberg 2001)
- 24.46 J. Walker: Rep. Prog. Phys. **42**, 1605–1654 (1979)
- 24.47 A. T. Collins, B. C. Lightowers: *The Properties of Diamond* (Academic, London 1979) p. 87
- 24.48 J. E. Field (Ed.): *The Properties of Natural and Synthetic Diamond* (Academic, New York 1992)
- 24.49 M. H. Nazaré, A. J. Neves (Eds.): *Properties, Growth and Applications of Diamond* (INSPEC, London 2001)
- 24.50 R. J. Trew, J. B. Yan, P. M. Mock: Proc. IEEE **79**, 598–620 (1991)
- 24.51 M. W. Geis, N. N. Efremow, D. D. Rathman: J. Vac. Sci. Technol. **A6**, 1953–1954 (1988)
- 24.52 K. Shenai, R. S. Scott, B. J. Baliga: IEEE Trans. Electron. Dev. **36**, 1811–1823 (1989)
- 24.53 B. J. Baliga: J. Appl. Phys. **53**, 1759 (1982)
- 24.54 A. M. Zaitsev: *Handbook of Diamond Technology* (Trans Tech, Zurich 2000) p. 198
- 24.55 S. Kumar, P. Ravindranathan, H. S. Dewan, R. Roy: Diamond Rel. Mater. **5**, 1246–1248 (1996)
- 24.56 Ch. Q. Sun, H. Xie, W. Zang, H. Ye, P. Hing: J. Phys. D: Appl. Phys. **33**, 2196–2199 (2000)
- 24.57 J. Chen, S. Z. Deng, J. Chen, Z. X. Yu, N. S. Su: Appl. Phys. Lett. **74**, 3651 (1999)
- 24.58 C. E. Jonhson, M. A. S. Hasting, W. A. Weimar: J. Mater. Res. **5**, 2320–2325 (1990)
- 24.59 Ch. Q. Sun, M. Alam: J. Electrochem. Soc. **139**, 933–936 (1992)
- 24.60 F. X. Lu, J. M. Liu, G. C. Chen, W. Z. Tang, C. M. Li, J. H. Song, Y. M. Tong: Diamond Rel. Mater. **13**, 533–538 (2004)
- 24.61 R. R. Nimmagadda, A. Joshi, W. L. Hsu: J. Mater. Res. **5**, 2445–2450 (1990)
- 24.62 F. K. de Theije, O. Roy, N. J. van der Laag, W. J. P. van Enckevort: Diamond Rel. Mater. **9**, 929–934 (2000)
- 24.63 F. K. de Theije, E. van Veenendaal, W. J. P. van Enckevort, E. Vlieg: Surface Sci. **492**, 91–105 (2001)
- 24.64 R. C. Burns, G. J. Davies: *The Properties of Natural and Synthetic Diamond* (Academic, London 1992)
- 24.65 B. V. Spitsyn, L. L. Bouilov, B. V. Derjaguin: J. Cryst. Growth **52**, 219–226 (1981)
- 24.66 H. Okushi: Diamond Rel. Mater. **10**, 281–288 (2001)
- 24.67 M. Kamo, Y. Sato, S. Matsumoto, N. J. Setaka: J. Cryst. Growth **62**, 642 (1983)
- 24.68 Y. Saito, S. Matsuda, S. Nogita: J. Mater. Sci. Lett. **5**, 565–568 (1986)
- 24.69 B. Dischler, C. Wild: *Low-Pressure Synthetic Diamond* (Springer, Berlin Heidelberg New York 1998)

- 24.70 P.W. May: Phil. Trans. R. Soc. Lond. **A 358**, 473 (2000)
- 24.71 J. Isberg, J. Hammarberg, E. Johansson, T. Wikström, D.J. Twitchen, A.J. Whitehead, S.E. Coe, G.A. Scarsbrook: Science **297**, 1670–1672 (2002)
- 24.72 J. Isberg, J. Hammarberg, D.J. Twitchen, A.J. Whitehead: Diamond Rel. Mater. **13**, 320–324 (2004)
- 24.73 B.R. Stoner, J.T. Glass: Appl. Phys. Lett. **60**, 698–700 (1992)
- 24.74 T. Tachibana, Y. Yokota, K. Nishimura, K. Miyata, K. Kobashi, Y. Shintani: Diamond Rel. Mater. **5**, 197–199 (1996)
- 24.75 H. Karawada, C. Wild, N. Herres, R. Locher, P. Koidl, H. Nagasawa: J. Appl. Phys. **81**, 3490–3493 (1997)
- 24.76 M. Oba, T. Sugino: Jpn. J. Appl. Phys. **39**, L1213–L1215 (2000)
- 24.77 S. Koizumi, T. Murakami, T. Inuzuka, K. Suzuki: Appl. Phys. Lett. **57**, 563–565 (1990)
- 24.78 S.D. Wolter, B.R. Stoner, J.T. Glass, P.J. Ellis, D.S. Buhaenko, C.E. Jenkins, P. Southworth: Appl. Phys. Lett. **62**, 1215–1217 (1993)
- 24.79 X. Jiang, C.P. Klages, R. Zachai, M. Hartweg, H.J. Füsser: Appl. Phys. Lett. **62**, 3438–3440 (1993)
- 24.80 C. Wild, R. Kohl, N. Herres, W. Müller-Sebert, P. Koidl: Diamond Rel. Mater. **3**, 373–381 (1994)
- 24.81 X. Jiang, C.P. Klages: *New Diamond and Diamond-Like Films* (Techna, Srl. 1995) pp. 23–30
- 24.82 S.D. Wolter, T.H. Borst, A. Vescan, E. Kohn: Appl. Phys. Lett. **68**, 3558 (1996)
- 24.83 A. Flöter, H. Güttler, G. Schulz, D. Steinbach, C. Lutz-Elsner, R. Zachai, A. Bergmaier, G. Dollinger: Diamond Rel. Mater. **7**, 283–288 (1998)
- 24.84 X. Jiang, K. Schifffmann, C.P. Klages, D. Wittorf, C.L. Jia, K. Urban, W. Jäger: J. Appl. Phys. **83**, 2511–2518 (1998)
- 24.85 P. Gluche, M. Adamschik, A. Vescan, W. Ebert, F. Szücs, H.J. Fecht, A. Flöter, R. Zachai, E. Kohn: Diamond Rel. Mater. **7**, 779–782 (1998)
- 24.86 P. Gluche, M. Adamschik: Diamond Rel. Mater. **8**, 934–940 (1999)
- 24.87 E. Kohn, A. Aleksov, A. Denisenko, P. Schmid, M. Adamschik, J. Kusterer, S. Ertl, K. Janischowsky, A. Flöter, W. Ebert: Diamond and Other Carbon Materials, Diamond in electronic applications, CIMTEC 2002 – 3rd Forum on New Materials 3rd International Conference, Florence 2002, ed. by P. Vincenzini, P. Ascarelli (Techna, 2003) 205–216
- 24.88 T. Tsubota, M. Ohta, K. Kusakabe, S. Morooka, M. Watanabe, H. Maeda: Diamond Rel. Mater. **9**, 1380–1387 (2000)
- 24.89 K. Ohtsuka, K. Suzuki, A. Sawabe, T. Inuzuka: Jpn. J. Appl. Phys. **35**, 1072 (1996)
- 24.90 M. Schreck, H. Roll, B. Stritzker: Appl. Phys. Lett. **74**, 650–652 (1999)
- 24.91 M. Schreck, A. Schury, F. Hörmann, H. Roll, B. Stritzger: J. Appl. Phys. **91**, 676–685 (2002)
- 24.92 M. Kubovic, A. Aleksov, M. Schreck, T. Bauer, B. Stritzker, E. Kohn: Diamond Rel. Mater. **12**, 403–407 (2003)
- 24.93 C. Bednarski, Z. Dai, A.P. Li, B. Golding: Diamond Rel. Mater. **12**, 241–245 (2003)
- 24.94 Z. Dai, C. Bednarski-Meinke, B. Golding: Diamond Rel. Mater. **13**, 552–556 (2004)
- 24.95 S. Gsell, T. Bauer, J. Goldfuss, M. Schreck, B. Stritzker: Appl. Phys. Lett. **84**, 4541–4543 (2004)
- 24.96 Infineon Technologies (2001), SDP04S60, SPD04S60, SDP06S60, SDB06S60, SDP10S30, SDB10S30, and SDB20S30 preliminary data sheets, homepage http://www.infineon.com/cgi/ecrm.dll/ecrm/scripts/search/advanced_search_result.jsp?queryString=SiC&x=22&y=7 (Jan. 2001)
- 24.97 M. Mazzola: *SiC high-temperatuer wideband gap mateials, Combat hybrid power system component technologies* (National Academic, Washington D.C. 2001) pp. 31–40
- 24.98 W. Wondrak, E. Niemann: Proc. IEEE International Symposium (ISIE '98), 1998, p. 153
- 24.99 T. Jang, L.M. Porter: Electrical characteristics of tantalum and tantalum carbide Schottky diodes on n- and p-type silicon carbide as a function of temperature, Proc. 4th International High Temperature Electronics Conference (HITEC 98) **4**, 280–286 (1998) Cat. No. 98EX145
- 24.100 Y. Sugawara, D. Takayama, K. Asano, R. Singh, J. Palmour, T. Hayashi: Int. Symp. on Power Semiconductor Devices & ICs, Osaka, Japan **13**, 27–30 (2001)
- 24.101 R. Singh, J.A.Jr. Cooper, R. Melloch, T.P. Chow, J. Palmour: IEEE Trans. Electron. Dev. **49**, 665 (2002)
- 24.102 R. Singh, K.G. Irvine, D. Cappel, T. James, A. Hefner, J.W. Palmour: IEEE Trans. Electron. Dev. **49**, 2308 (2002)
- 24.103 R. Singh, D.C. Capell, K.G. Irvine, J.T. Richmond, J.W. Palmour: Electron. Lett. **38**, 1738 (2002)
- 24.104 B.M. Wilamowski: Solid State Electron. **26**, 491 (1983)
- 24.105 B.J. Baliga: IEEE Elect. Dev. Lett. **5**, 194 (1984)
- 24.106 C.M. Zetterling, F. Dahlquist, N. Lundberg, M. Ostling: Solid State Electron. **42**, 1757 (1998)
- 24.107 F. Dahlquist, C.M. Zetterling, M. Ostling, K. Rutner: Mater. Sci. Forum **264–268**, 1061 (1998)
- 24.108 F. Dahlquist, H. Lendenmann, M. Ostling: Mater. Sci. Forum **389–393**, 1129 (2002)
- 24.109 D. Nakamura, I. Gunjishima, S. Yamagushi, T. Ito, A. Okamoto, H. Kondo, S. Onda, K. Takatori: Nature **430**, 1009–1012 (2004)
- 24.110 R.C. Clarke, A.W. Morse, P. Esker, W.R. Cur-tice: Proc. IEEE High Performance Devices, 2000 IEEE/Cornell Conference, 141–143 (2000)
- 24.111 S.T. Sheppard, R.P. Smith, W.L. Pribble, Z. Ring, T. Smith, S.T. Allen, J. Milligan, J.W. Palmour: Proc. Device Research Conference, 60th DRC. Conference Digest, 175–178 (2002)

- 24.112 S. T. Allen, W. L. Pribble, R. A. Sadler, T. S. Alcorn, Z. Ring, J. W. Palmour: IEEE MTT-S Int. **1**, 321–324 (1999)
- 24.113 F. Villard, J. P. Prigent, E. Morvan, C. Brylinski, F. Temcamani, P. Pouvil: IEEE Trans. Microw. Theory Technol. **51**, 1129 (2003)
- 24.114 J. B. Casady, E. D. Luckowski, R. W. Johnsson, J. Crofton, J. R. Williams: IEEE Electronic Components and Technology Conference **45**, 261–265 (1995)
- 24.115 A. Agarwal, C. Capell, B. Phan, J. Miligan, J. W. Palmour, J. Stambaugh, H. Bartlow, K. Brewer: IEEE High Performance Devices, 2002, Proceedings, IEEE Lester Eastman Conference, p. 41
- 24.116 B. J. Baliga: Proc. 6th Int. Conf. on SiC & Related Mat. Ser. 142, 1–6 (1996)
- 24.117 K. Asano, Y. Sugawara, T. Hayashi, S. Ryu, R. Singh, J. W. Palmour, D. Takayama: IEEE Power Semiconductor Devices and ICs, 2002. Proceedings of the 14th International Symposium, p. 61
- 24.118 J. H. Zhao, X. Li, K. Tone, P. Alexandrov, M. Pan, M. Weiner: IEEE Semiconductor Device Research Symposium, 2001 International, p. 564
- 24.119 C. J. Scozzie, C. Wesley, J. M. McGarrity, F. B. Mclean: Reliability Physics Symposium, 1994. 32nd Annual Proceedings., IEEE International, p. 351
- 24.120 K. Asano, Y. Sugawara, S. Ryu, J. W. Palmour, T. Hayashi, D. Takayama: Proc. 13th Int. Symp. Power Semiconductor Devices and ICs, Osaka, Japan, 2001
- 24.121 N. S. Rebello, F. S. Shoucair, J. W. Palmour: IEEE Proc. Circuits Dev. Syst. **143**, 115 (1996)
- 24.122 T. Billon, T. Ouisse, P. Lassagne, C. Jassaud, J. L. Ponthenier, L. Baud, N. Becourt, P. Morfouli: Electron. Lett. **30**, 170 (1994)
- 24.123 S. H. Ryu, A. Agarwal, J. Richmond, J. Palmour, N. Saks, J. Williams: IEEE-Power Semiconductor Devices and ICs, 2002. Proceedings of the 14th International Symposium, 65–68 (2002)
- 24.124 J. B. Casady, A. Agarwal, L. B. Rowland, S. Shshadri, R. R. Siergiej, D. C. Sheridan, S. Mani, P. A. Sanger, C. D. Brandt: IEEE-Compound Semiconductors, International Symposium, 1998, p. 359
- 24.125 J. Spitz, M. R. Melloch, J. A. Jr. Cooper, M. A. Capano: IEEE Electron. Dev. Lett. **19**, 100 (1998)
- 24.126 I. A. Khan, J. A. Jr. Cooper, M. Capano, T. Isaacs-Smith, J. R. Williams: IEEE – Power Semiconductor Devices and ICs, [2002] Proceedings of the 14th International Symposium, 157–160 (2002)
- 24.127 R. K. Chilukuri, M. Praveen, B. J. Baliga: IEEE Trans. Ind. Appl. **35**, 1458 (1999)
- 24.128 G. I. Gudjonsson, H. O. Olafsson, E. O. Sveinbjornsson: Mater. Sci. Forum **457–460**, 1425–1428 (2004)
- 24.129 H. O. Olafsson, G. I. Gudjonsson, P. O. Nillson, E. O. Sveinbjornsson, H. Zirath, R. Rodle, R. Jos: Electron. Lett. **40**, 508–509 (2004)
- 24.130 L. A. Lipkin, J. W. Palmour: Mater. Sci. Forum **338**, 1093 (2000)
- 24.131 R. C. Clarke, C. D. Brandt, S. Sriram, R. R. Siergiej, A. W. Morse, A. Agarwal, L. S. Chen, V. Balakrishna, A. A. Burk: Proc. IEEE High Temperature Electronic Materials, Devices and Sensors Conference, 1998, p. 18
- 24.132 S. H. Ryu, A. Agarwal, R. Singh, J. W. Palmour: IEEE Trans. Electron. Dev. Lett. **22**, 124 (2001)
- 24.133 S. F. Kozlov: IEEE Trans. Nucl. Sci **222**, 160 (1975)
- 24.134 S. F. Kozlov, E. A. Komorova, Y. A. Kuznetsov, Y. A. Salikov, V. I. Redko, V. R. Grinberg, M. L. Meilman: IEEE Trans. Nucl. Sci **24**, 235–237 (1977)
- 24.135 M. Krammer, W. Adam, E. Berdermann, P. Bergonzo, G. Bertuccio, F. Bogani, E. Borch, A. Brambilla, M. Bruzzi, C. Colledani: Diamond Rel. Mater. **10**, 1778–1782 (2001)
- 24.136 A. Mainwood: Semicond. Sci. Technol. **15**, 55 (2000)
- 24.137 M. Adamschik, M. Müller, P. Gluche, A. Flöter, W. Limmer, R. Sauer, E. Kohn: Diamond Rel. Mater. **10**, 1670–1675 (2001)
- 24.138 A. Denisenko, A. Aleksov, E. Kohn: Diamond Rel. Mater. **10**, 667–672 (2001)
- 24.139 M. D. Whitfield, S. P. Lansley, O. Gaudin, R. D. McKeag, N. Rizvi, R. B. Jackman: Diamond Rel. Mater. **10**, 715–721 (2001)
- 24.140 J. Gondolek, J. Kocol: Diamond Rel. Mater. **10**, 1511–1514 (2001)
- 24.141 P. R. Chalker, T. B. Joyce, C. Johnston: Diamond Rel. Mater. **8**, 309–313 (1999)
- 24.142 E. Kohn, M. Adamschik, P. Schmid, S. Ertl, A. Flöter: Diamond Rel. Mater. **10**, 1684–1691 (2001)
- 24.143 T. H. Borst, O. Weis: Diamond Rel. Mater. **4**, 948–953 (1995)
- 24.144 B. A. Fox, M. L. Hartsell, D. M. Malta, H. A. Wynands, G. J. Tessmer, D. L. Dreifus: Mater. Res. Soc. Symp. Proc. **416**, 319 (1996)
- 24.145 S. Yamanaka, H. Watanabe, S. Masai, D. Takeuchi, H. Okushi, K. Kajimura: Jpn. J. Appl. Phys. **37**, 1129 (1998)
- 24.146 D. Saito, E. Tsutsumi, N. Ishigaki, T. Tashiro, T. Kimura, S. Yugo: Diamond Relat. Mater. **11**, 1804–1807 (2002)
- 24.147 A. Aleksov, A. Denisenko, M. Kunze, A. Vescan, A. Bergmeir, G. Dollinger, W. Ebert, E. Kohn: Semicond. Sci. Technol. **18**, S59–S66 (2003)
- 24.148 M. I. Landstrass, K. V. Ravi: Appl. Phys. Lett. **55**, 975–977 (1989)
- 24.149 M. I. Landstrass, K. V. Ravi: Appl. Phys. Lett. **55**, 1391–1393 (1989)
- 24.150 S. Koizumi, K. Watanabe, M. Hasegawa, H. Kanda: Diamond Rel. Mater. **11**, 307–311 (2002)
- 24.151 M. Katagiri, J. Isoya, S. Koizumi, H. Kanda: Appl. Phys. Lett. **85**, 6365–6367 (2004)
- 24.152 E. Gheeraerf, A. Casanova, A. Tajani, A. Deneuville, E. Bustarret, J. A. Garrido, C. E. Nebel, M. Stutzmann: Diamond Rel. Mater. **11**, 289–295 (2000)

- 24.153 J. F. Prins: *Diamond Rel. Mater.* **10**, 1756–1764 (2001)
- 24.154 H. Sternschulte, M. Schreck, B. Stritzker, A. Bergmaier, G. Dollinger: *Diamond Rel. Mater.* **9**, 1046–1050 (2000)
- 24.155 R. Kalish: *Diamond Rel. Mater.* **10**, 1749–1755 (2001)
- 24.156 A. Aleksov, A. Denisenko, E. Kohn: *Solid-State Electron.* **44**, 369–375 (2000)
- 24.157 P. Gluche, A. Vescan, W. Ebert, M. Pitter, E. Kohn: Processing of High Temperature Stable Contacts to Single Crystal Diamond, *Transient Thermal Processing Techniques in Electronic Materials*, Proc. 12th TMS Annual Meeting, Anaheim 1996, ed. by N. M. Ravindra, R. K. Singh (Warrendale, PA 1996) 107–110
- 24.158 A. Vescan, I. Daumiller, P. Gluche, W. Ebert, E. Kohn: *IEEE Electron. Dev. Lett.* **18**, 556–558 (1997)
- 24.159 A. Vescan, I. Daumiller, P. Gluche, W. Ebert, E. Kohn: *Diamond Relat. Mater.* **7**, 581–584 (1998)
- 24.160 J. A. Garrido, C. E. Nebel, R. Todt, G. Rösel, M.-C. Amann, M. Stutzmann, E. Snidero, P. Bergonzo: *Appl. Phys. Lett.* **82**, 988–990 (2003)
- 24.161 T. Maki, S. Shikama, M. Komori, Y. Sakagushi, K. Sakura, T. Kobayashi: *Jpn. J. Appl. Phys.* **31**, 1446–1449 (1992)
- 24.162 H. Karawada: *Surface Sci. Rep.* **26**, 205–259 (1996)
- 24.163 K. Hayashi, S. Yamanaka, H. Okushi, K. Kajimura: *Appl. Phys. Lett.* **68**, 376–378 (1996)
- 24.164 K. Hayashi, S. Yamanaka, H. Watanabe, T. Sekiguchi, H. Okushi, K. Kajimura: *J. Appl. Phys.* **81**, 744–753 (1997)
- 24.165 J. Ristein, F. Maier, M. Riedel, J. B. Cui, L. Ley: *Phys. Status Solidi A* **181**, 65 (2000)
- 24.166 F. Maier, M. Riedel, B. Mantel, J. Ristein, L. Ley: *Phys. Rev. Lett.* **85**, 3472–3475 (2000)
- 24.167 B. F. Mantel, M. Stammer, J. Ristein, L. Ley: *Diamond Rel. Mater.* **10**, 429–433 (2001)
- 24.168 J. Ristein, F. Maier, M. Riedel, M. Stammer, L. Ley: *Diamond Rel. Mater.* **10**, 416–422 (2001)
- 24.169 T. Yamada, A. Kojima, A. Sawaba, K. Suzuki: Passivation of hydrogen terminated diamond surface conductive layer using hydrogenated amorphous carbon, *Diamond Rel. Mater.* **13**, 776–779 (2004)
- 24.170 Y. Otsuka, S. Suzuki, S. Shikama, T. Maki, T. Kobayashi: *Jpn. J. Appl. Phys.* **34**, 551 (1995)
- 24.171 O. A. Williams, R. B. Jackman, C. Nebel, J. S. Foord: *Diamond Rel. Mater.* **11**, 396–399 (2002)
- 24.172 B. A. Fox, M. L. Hartsell, D. M. Malta, H. A. Wynands, C.-T. Kao, L. S. Plano, G. J. Tessmer, R. B. Henard, J. S. Holmes, A. J. Tessmer, D. L. Dreifus: *Diamond Rel. Mater.* **4**, 622–627 (1995)
- 24.173 K. Tsugawa, H. Kitatani, A. Noda, A. Hokazono, K. Hirose, M. Tajima: *Diamond Rel. Mater.* **8**, 927–933 (1999)
- 24.174 A. Aleksov, M. Kubovic, N. Kaeb, U. Spitzberg, A. Bergmaier, G. Dollinger, Th. Bauer, M. Schreck, B. Stritzker, E. Kohn: *Diamond Rel. Mat.* **12**, 391–398 (2003)
- 24.175 H. Ishizaka, H. Umezawa, H. Taniuchi, T. Arima, N. Fujihara, M. Tachiki, H. Kawarada: *Diamond Rel. Mat.* **11**, 378–381 (2002)
- 24.176 P. Gluche, A. Aleksov, V. Vescan, W. Ebert, E. Kohn: *IEEE Electron. Dev. Lett.* **18**, 547–549 (1997)
- 24.177 H. Taniuchi, H. Umezawa, T. Arima, M. Tachiki, H. Kawarada: *IEEE Electron. Dev. Lett.* **22**, 390–392 (2001)
- 24.178 A. Aleksov, A. Denisenko, U. Spitzberg, T. Jenkins, W. Ebert, E. Kohn: *Diamond Rel. Mat.* **11**, 382–386 (2002)
- 24.179 A. Aleksov, M. Kubovic, M. Kasu, P. Schmid, D. Grobe, S. Ertl, M. Schreck, B. Stritzker, E. Kohn: *Diamond Rel. Mater.* **13**, 233–240 (2004)
- 24.180 M. Kasu, M. Kubovic, A. Aleksov, N. Teofilov, Y. Taniyasu, R. Sauer, E. Kohn, T. Makimoto, N. Kobayashi: *Diamond Rel. Mater.* **13**, 226–232 (2004)
- 24.181 M. Kunze, A. Vescan, G. Dollinger, A. Bergmaier, E. Kohn: *Carbon* **37**, 787–791 (1999)
- 24.182 A. Aleksov, A. Vescan, M. Kunze, P. Gluche, W. Ebert, E. Kohn, A. Bergmeier, G. Dollinger: *Diamond Rel. Mater.* **8**, 941–945 (1999)
- 24.183 A. Vescan, P. Gluche, W. Ebert, E. Kohn: *IEEE Electron Dev. Lett.* **18**, 222–224 (1997)
- 24.184 W. P. Kang, T. S. Fischer, J. L. Davidson: *New Diamond Frontier Carbon Technol.* **11**, 129 (2001)
- 24.185 J. L. Davidson, W. P. Kang, A. Wisitsora-At: *Diamond Rel. Mater.* **12**, 429–433 (2003)
- 24.186 D. J. Twitchen, C. S. J. Pickles, S. E. Coe, R. S. Sussmann, C. E. Hall: *Diamond Rel. Mater.* **10**, 731–735 (2001)
- 24.187 W. Ebert, M. Adamschik, P. Gluche, A. Flöter, E. Kohn: *Diamond Rel. Mater.* **8**, 1875–1877 (1999)
- 24.188 E. Kohn, M. Adamschik, P. Schmid, S. Ertl, A. Flöter: *Diamond Rel. Mater.* **10**, 1684–1691 (2001)
- 24.189 E. Worner: *Low Pressure Synthetic Diamond* (Springer, Berlin Heidelberg New York 1998) p. 137
- 24.190 M. Seelmann-Eggebert, P. Meisen, F. Schaudel, P. Koidl, A. Vescan, H. Leier: *Diamond Rel. Mater.* **10**, 744–749 (2001)
- 24.191 Y. F. Wu, B. P. Keller, S. Keller: *IEICE Trans. Electron.* **E82(11)**, 1895 (1999)
- 24.192 C. E. Weitzel: *IOP Conf. Series* **142**, 765 (1996)
- 24.193 B. C. Djubua, N. N. Chubun: *IEEE Trans. Electron. Dev.* **38**, 2314 (1991)
- 24.194 J. Robertson: *J. Vac. Sci. Technol.* **B 17**, 659–665 (1999)
- 24.195 J. B. Cui, M. Stammler, J. Ristein, L. Ley: *J. Appl. Phys.* **88**, 3667–3673 (2000)
- 24.196 W. Choi, Y. D. Kim, Y. Iseri, N. Nomura, H. Tomokage: *Diamond Rel. Mat.* **10**, 863–867 (2001)
- 24.197 K. M. Song, J. Y. Shim, H. K. Baik: *Diamond Rel. Mat.* **11**, 185–190 (2002)
- 24.198 P. H. Cutler, N. M. Miskovsky, P. B. Lerner, M. S. Ching: *Appl. Surf. Sci.* **146**, 126–133 (1999)

- 24.199 X.-Z. Ding, B. K. Tay, S. P. Lau, J. R. Shi, Y. J. Li, Z. Sun, X. Shi, H. S. Tan: *J. Appl. Phys.* **88**, 5087–5092 (2000)
- 24.200 J. van der Weide, R. J. Nemanich: *J. Vac. Sci. Technol.* **B 10**, 1940–1943 (1992)
- 24.201 T. S. Fischer, D. G. Walker: *J. Heat Trans.* **124**, 954–962 (2002)
- 24.202 F. A. M. Köck, J. M. Garguilo, B. Brown, R. J. Nemanich: *Diamond Rel. Mater.* **11**, 774 (2002)
- 24.203 S. H. Shin, T. S. Fisher, D. G. Walker, A. M. Strauss, W. P. Kang, J. L. Davidson: *J. Vac. Sci. Technol.* **B 21**, 587–592 (2003)
- 24.204 J. L. Davidson, W. Kang, K. Holmes, A. Wisitsorath, P. Taylor, V. Pulugurta, R. Venkatasubramanian, F. Wells: *Diamond Rel. Mater.* **10**, 1736–1742 (2001)
- 24.205 V. K. Dimitrev, V. N. Inkin, G. G. Kirpilenko, B. G. Potapov, E. A. Ylyichev, E. Y. Shelukhin: *Diamond Rel. Mater.* **10**, 1007–1010 (2001)

25. Amorphous Semiconductors: Structure, Optical, and Electrical Properties

This chapter is devoted to a survey of the structural, optical and electrical properties of amorphous semiconductors on the basis of their fundamental understanding. These properties are important for various types of applications using amorphous semiconductors.

First, we review general aspects of the electronic states and defects in amorphous semiconductors, i.e., a-Si:H and related materials, and chalcogenide glasses, and their structural, optical and electrical properties.

Further, we survey the two types of phenomena associated with amorphous structure, i.e., light-induced phenomena, and quantum phenomena associated with nanosized amorphous structure. The former are important from the viewpoint of amorphous-silicon solar cells. The latter phenomena promise novel applications of amorphous semiconductors from the viewpoint of nanotechnology.

25.1	Electronic States	565
25.2	Structural Properties	568
25.2.1	General Aspects	568
25.2.2	a-Si:H and Related Materials	568
25.2.3	Chalcogenide Glasses	569
25.3	Optical Properties	570
25.3.1	General Aspects	570
25.3.2	a-Si:H and Related Materials	571
25.3.3	Chalcogenide Glasses	572
25.4	Electrical Properties	573
25.4.1	General Aspects	573
25.4.2	a-Si:H and Related Materials	574
25.4.3	Chalcogenide Glasses	575
25.5	Light-Induced Phenomena	575
25.6	Nanosized Amorphous Structure	577
	References	578

Amorphous semiconductors are promising electronic materials for a wide range of applications such as solar cells, thin-film transistors, light sensors, optical memory devices, vidicons, electrophotographic applications, X-ray image sensors, europium-doped optical-fibre amplifications etc, particularly, hydrogenated amorphous silicon (a-Si:H) for solar cells, thin-film transistors, X-ray image sensors, and chalcogenide glasses for optical memory devices including digital video/versatile disk (DVD). In this chapter, we emphasize the basic concepts and general aspects of the electronic properties of amorphous semiconductors such as their electrical and optical properties as well as their struc-

tural properties [25.1–3]. Furthermore, some basic and important results of these properties are described to understand these applications and to consider their further development. Light-induced phenomena in amorphous semiconductors, which have been considered to be associated with amorphous structure, are also described.

Nanosized amorphous structures exhibit quantum effects associated with two-dimensional (quantum well), one-dimensional (quantum wire) and zero-dimensional (quantum dot) structures, so they have received significant attention on both the fundamental and application sides. These topics are briefly described.

25.1 Electronic States

Long-range disorder in amorphous network breaks down the periodic arrangement of constituent atoms, as shown in Fig. 25.1. In the figure, the structures of amorphous

and crystalline silicon are shown. The periodic arrangement of atoms makes it easy to treat the electronic states mathematically, i.e., the so-called Bloch the-

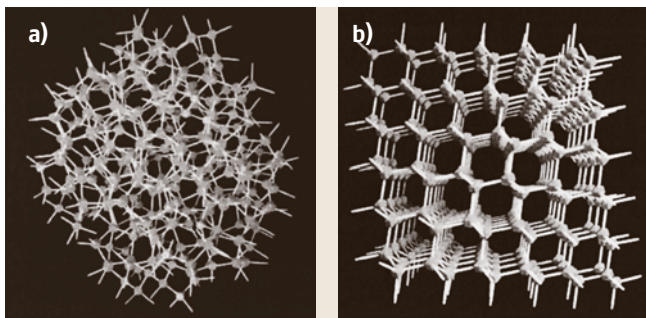


Fig. 25.1a,b Structural models of (a) amorphous silicon and (b) crystalline silicon

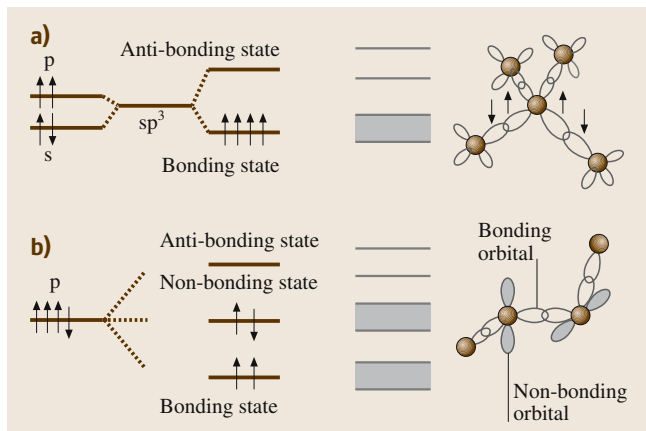


Fig. 25.2a,b Schematic diagram of the energy levels of atomic orbitals, hybridized orbitals and bands for (a) tetrahedrally bonded semiconductors, e.g., a-Si and for (b) selenium

ory can be applied to crystalline solids. On the other hand, it becomes difficult to treat the electronic states in amorphous solids mathematically. However, electronic states in amorphous semiconductors are simply described in terms of tight-binding approximations and Hartree–Fock calculations. Using these approaches, spatial fluctuations of bond length, bond angle and dihedral angle lead to broadening of the edges of the conduction and valence bands, constructing the band tail. This is due to spatial fluctuations of bond energy between constituent atoms in tetrahedrally coordinated semiconductors such as Si and Ge, in which the bonding state and the antibonding state constitute the valence band and the conduction band, respectively, as shown in Fig. 25.2a. In chalcogenide glasses, the conduction band arises from the antibonding state, while the valence band arises from the nonbonding state, as shown in Fig. 25.2b. The broadening of those bonds also occurs as a result of potential

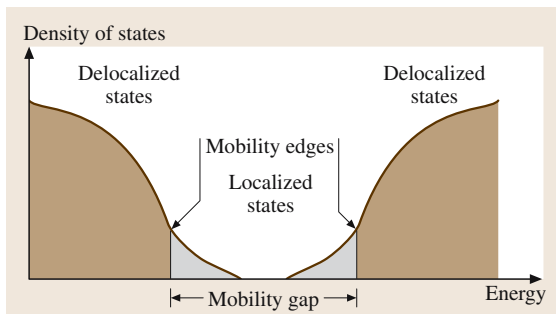


Fig. 25.3 Schematic diagram of the density of states of the conduction band (delocalized states) and the valence band (delocalized states) shown by the brown-colored. The mobility edge and mobility gap are shown

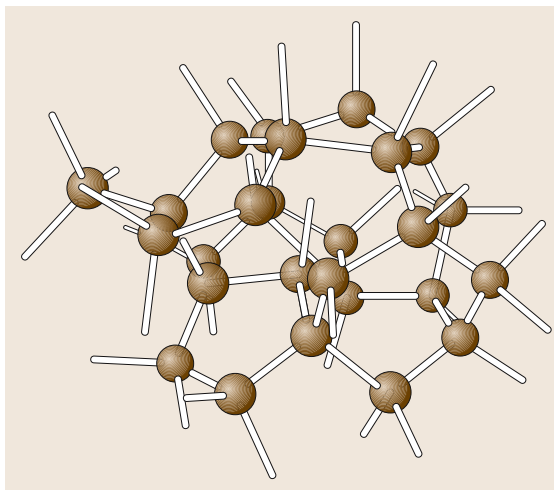


Fig. 25.4 Schematic illustration of a dangling bond in tetrahedrally bonded amorphous semiconductors

fluctuations associated with the amorphous network of constituent atoms. The static charge fluctuation associated with bond-length and bond-angle variations in the amorphous network has been discussed theoretically [25.4] and experimentally [25.5, 6] in amorphous silicon.

In the band tails, the electronic states have a localized character and their nature changes from localized to delocalized at a critical boundary called the mobility edge, as shown in Fig. 25.3. The energy separation between the two mobility edges of the conduction and valence bands is called the mobility gap. The nature of the conduction and valence bands has been elucidated by means of photoemission spectroscopy [25.7]. The structure of actual samples of amorphous semiconduc-

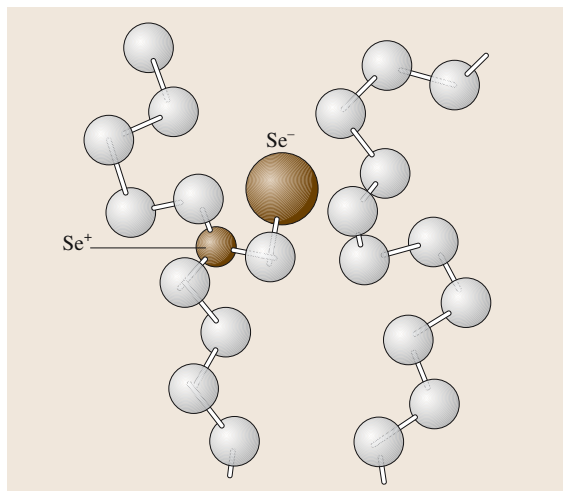


Fig. 25.5 Schematic illustration of two selenium chains with positively charged threefold-coordinated selenium and negatively charged onefold-coordinated selenium

tors deviates from the ideal random network, namely, coordination of constituent atoms deviates from the normal coordination following the $8 - N$ rule [25.1], where N designates the number of valence electrons. For instance, the normal coordinations of Si and Se are four and two, respectively. However, in actual samples Si atoms with a threefold coordination are present in a-Si (Fig. 25.4), and Se atoms with onefold and threefold coordinations are present in a-Se (Fig. 25.5). These atoms are generally called structural defects (gap states) whose electronic energy levels are located within the band-gap region. In the case of chalcogenide glasses, onefold- and threefold-coordinated Se atoms are called valence-alternation pairs (VAP) [25.8]. The detailed properties of structural defects have been elucidated by means of optical spectroscopies such as photoinduced absorption (PA), photothermal deflection spectroscopy (PDS) and the constant-photocurrent method (CPM), capacitance measurements such as deep-level transient spectroscopy (DLTS) and isothermal capacitance transient spectroscopy (ICTS), and resonance methods such as electron spin resonance (ESR) and electron–nuclear double resonance (ENDOR) [25.2].

In the following, we take a-Si:H as an example and describe the nature of its electronic states. Figure 25.6 shows a schematic diagram of the electronic states involved in the band-gap region of a-Si:H proposed by the authors' group [25.9]. The tails of the

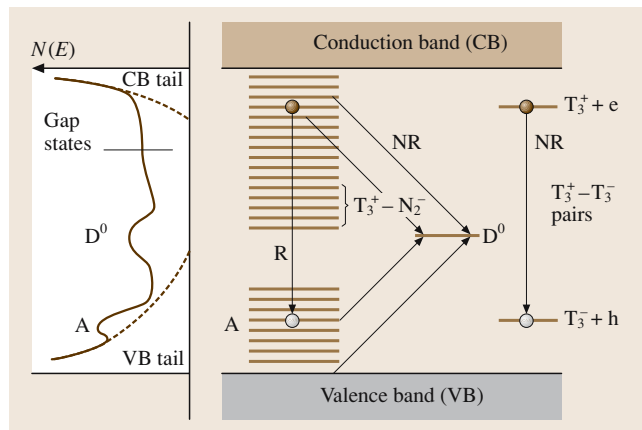


Fig. 25.6 Schematic diagram of the tail and gap states and the recombination processes in a-Si:H. R and NR designate radiative and nonradiative recombination, respectively. A diagram of the density of states spectrum is illustrated on the left-hand side. For the definition of other symbols, see the text. (After [25.9])

conduction and valence bands are expressed by an exponential function with widths of 25 meV [25.10] and 48–51 meV [25.11–13], respectively.

In this figure, the processes of radiative recombination and nonradiative recombination are shown, along with the energy levels of the gap states involved in the recombination processes. The radiative recombination occurs between tail electrons and tail holes at higher temperatures and between tail electrons and self-trapped holes (A centres) at lower temperatures, particularly below 60 K. Such recombinations contribute to the principal luminescence band peaked at 1.3–1.4 eV. These have been elucidated by optically detected magnetic resonance (ODMR) measurements [25.14–16], and are briefly mentioned in Sect. 25.3. The radiative recombination center participating in the low-energy luminescence (defect luminescence) may be attributed to distant $T_3^+ - N_2^-$ pair defects [25.17, 18], in which T_3^+ and N_2^- are the positively charged threefold-coordinated silicon center and the negatively charged twofold-coordinated nitrogen center, respectively. The N_2^- centers are created from contaminating nitrogen atoms introduced during sample preparation. On the other hand, close $T_3^+ - N_2^-$ pair defects act as nonradiative recombination centers. A typical nonradiative recombination center in a-Si:H is the neutral silicon dangling bond, i. e., T_3^0 . The details of the recombination processes in a-Si:H are described in Sect. 25.3.

25.2 Structural Properties

25.2.1 General Aspects

The structural properties of amorphous semiconductors have been investigated by means of X-ray, electron and neutron diffraction, transmission electron microscope (TEM) and scanning electron microscope (SEM), extended X-ray absorption fine structure (EXAFS), small-angle X-ray scattering (SAXS), Raman scattering, infrared absorption (IR) and nuclear magnetic resonance (NMR). The absence of long-range order in amorphous semiconductors is manifested in diffraction techniques, e.g., electron diffraction for a-Si results in a halo pattern as shown in Fig. 25.7a, while Laue spots are seen for c-Si, as shown in Fig. 25.7b. The medium-range order of 0.5–50 nm has been discussed based on SAXS measurements.

25.2.2 a-Si:H and Related Materials

The radial distribution function (RDF) in a-Si has been obtained from an analysis of curves of scattered electron intensity versus scattering angle from electron diffraction measurements, as shown in Fig. 25.8 [25.19]. From a comparison of RDF curves of a-Si and c-Si, it was concluded that the first peak of a-Si coincides with that of c-Si, the second peak of a-Si is broadened compared to that of c-Si, the third peak of c-Si almost disappears and a small peak appears, shifted from the third peak of c-Si, as shown in Fig. 25.8. These results indicate that the bond length of a-Si is elongated by 1% over that of c-Si ($= 2.35 \text{ \AA}$), and that the bond angle is tetrahedral ($= 109.47^\circ$) with a variation of 10% from that of c-Si. Using the neutron diffraction technique, the RDF was also deduced from the measured structure factor, $S(Q)$, as a function of Q (the momentum transfer) in a-Si [25.20]. The RDF curve of a-Si exhibiting these features was well simulated by

the continuous random network model [25.21]. Since then, computer-generated models have been constructed (see, e.g., Kugler et al. [25.20]), using the Monte Carlo (MC) method [25.22], the molecular dynamics (MD) method [25.23] and the reverse MC method [25.24, 25]. Here, the reverse MC method, i.e., a MC simulation repeatedly carried out to reach consistency with the measured RDF curve and curve of structure factor $S(Q)$ versus Q from the neutron diffraction data, is briefly mentioned. First we start with an initial set of Cartesian coordinates (particle configuration) and calculate its RDF and $S(Q)$. Comparing the calculated RDF and $S(Q)$ with the measured ones, the new particle configuration is generated by random motion of a particle, being consistent with the constraint, e.g., the coordination number. From the reverse MC simulation, the bond-angle distribution is derived, i.e., it shows a peak at the tetrahedral angle, 109.5° and small peaks at $\approx 60^\circ$ and $\approx 90^\circ$ [25.26].

In a-Si:H, knowledge of the hydrogen configuration is important to understand its electronic properties. This is obtained by NMR [25.27, 28] and IR measurements [25.29]. The bonding modes of Si and H atoms are Si–H, Si–H₂, (Si–H₂)_n and Si–H₃ bonds. The concentrations of these hydrogen configurations depend on the preparation conditions, particularly on the substrate temperature, i.e., the deposition temperature. a-Si:H films prepared at 250 °C mostly contain Si–H bonds in incorporated hydrogen (hydrogen content of $\approx 10 \text{ at. \%}$), while those prepared at lower temperatures such as room temperature contain all these bonds in incorporated hydrogen (hydrogen content of

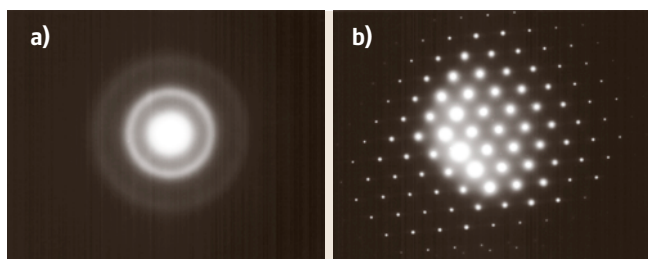


Fig. 25.7a,b Electron diffraction patterns. (a) amorphous silicon, (b) crystalline silicon

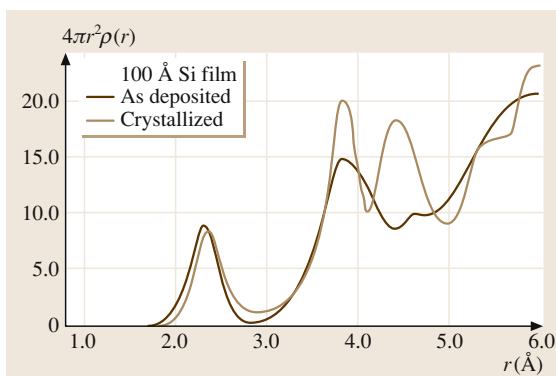


Fig. 25.8 Radial distribution function of amorphous (evaporated) and crystalline silicon obtained from electron diffraction patterns. (After [25.19])

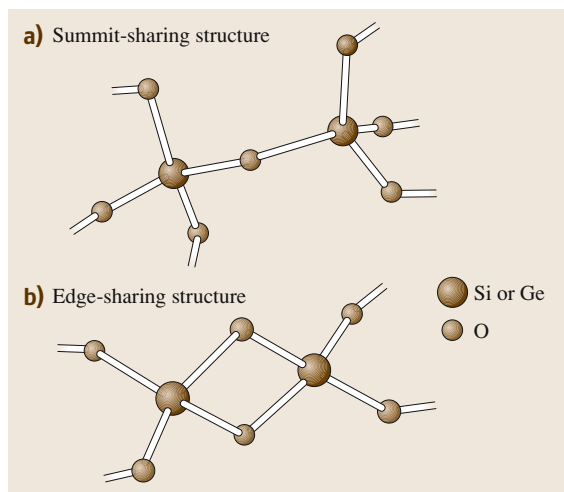


Fig. 25.9a,b Schematic illustration of (a) summit-sharing structure and (b) edge-sharing structure. (After [25.32])

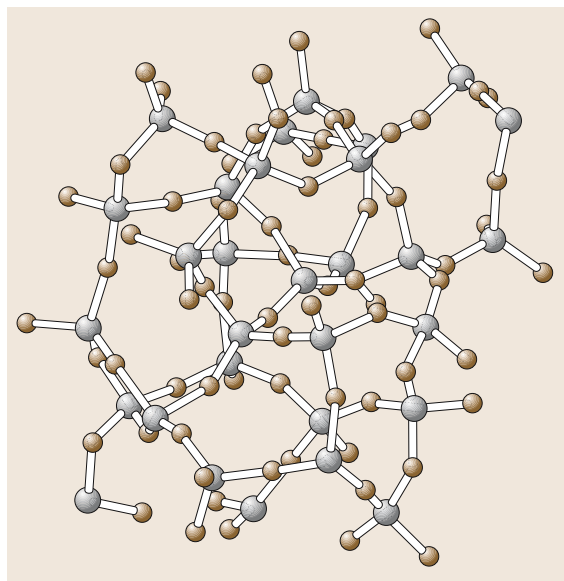


Fig. 25.10 Structural model of germanium disulfide. (After [25.32])

≈ 30 at. %). The hydrogen configuration has also been investigated by neutron scattering measurements using the isotope-substitution method [25.30,31]. Partial pair-correlation functions of Si–Si, Si–H and H–H are

obtained from those measurements. The Si–H bond length is 1.48 Å.

Raman scattering measurements provide useful information about microcrystallinity, i.e., the volume fraction of the amorphous phase in microcrystalline silicon [25.33]. The Raman shift associated with transverse optical (TO) phonons is observed at 520 cm^{-1} for c-Si, while it is observed at 480 cm^{-1} for a-Si.

25.2.3 Chalcogenide Glasses

The chalcogenide glasses are composed of chalcogenide atoms and other constituent atoms such as As, Ge etc. Their crystalline counterpart has a structure consisting of threefold-coordinated As atoms and twofold-coordinated chalcogens, i.e., c-Se and c-As₂S₃ exhibit a chain structure and a layered structure, respectively. The structure of the chalcogenide glasses is fundamentally built from these constituent atoms, keeping their coordination in the crystalline counterpart. However, actual glasses contain wrong bonds, i.e., homopolar bonds such as Se–Se, As–As, Ge–Ge etc. In the following, we take two examples of chalcogenide glasses, As₂X₃ ($X = \text{S, Se}$) and GeX₂ ($X = \text{S, Se}$).

1. As₂X₃ ($X = \text{S, Se}$)

This type of glass, e.g., As₂S₃, is mainly composed of AsS₃ pyramid units, i.e., a random network of these units sharing the twofold-coordinated S site [25.34]. There is no correlation between As atoms, but interlayer correlation seems to hold even for the amorphous structure.

2. GeX₂ ($X = \text{S, Se}$)

Silica glass (SiO₂) is a well-known material, which is mainly composed of SiO₄ tetrahedron units sharing the twofold-coordinated O site (the summit-sharing structure. See Fig. 25.9a). On the other hand, SiS₂ and GeSe₂ glasses have a different structure from SiO₂ glass: the tetrahedron units are connected through two X sites (the edge-sharing structure) along with one X site (the summit-sharing structure. See Fig. 25.9b). For example, in GeSe₂, the edge-sharing structure occurs at a level of more than 30% and the summit-sharing structure with less than 70%. Actually, wrong bonds (homopolar bonds) Ge–Ge, Se–Se exist in GeSe₂ glass. A structural model of GeS₂ [25.32] is shown in Fig. 25.10.

25.3 Optical Properties

25.3.1 General Aspects

Optical absorption and luminescence occur by transition of electrons and holes between electronic states such as conduction and valence bands, tail states, and gap states. In some cases, electron–phonon coupling is strong and, as a result, self-trapping occurs. Exciton formation has also been suggested in some amorphous semiconductors.

The absorption of photons due to interband transition, which occurs in crystalline semiconductors, is also observed in amorphous semiconductors. However, the absorption edge is not clear since interband absorption near the band gap is difficult to distinguish from tail absorption in the absorption spectra. Figure 25.11 schematically illustrates typical absorption spectra of amorphous semiconductors. The absorption coefficient, α , due to interband transition near the band gap is known to be well described by the following equation [25.35]

$$\alpha \hbar\omega = B(\hbar\omega - E_g)^2, \quad (25.1)$$

where $\hbar\omega$ and E_g denote the photon energy and optical gap, respectively. In most amorphous semiconductors, the optical gap E_g is determined by a plot of $(\alpha \hbar\omega)^{1/2}$ versus $\hbar\omega$, which is known as Tauc's plot. The photon energy at which the absorption coefficient is 10^4 cm^{-1} , E_{04} , is also used for the band gap in a-Si:H. The absorption coefficient at the photon energy just below the optical gap (tail absorption) depends exponentially on the photon energy, $E = \hbar\omega$, as expressed by

$$\alpha(E) \propto \exp\left(\frac{E}{E_U}\right), \quad (25.2)$$

where E_U is called the Urbach energy and also the Urbach tail width. In addition, optical absorption by defects appears at energies lower than the optical gap.

Photoluminescence (PL) occurs as a result of the transition of electrons and holes from excited states to the ground state. After interband excitation, electrons relax to the bottom of the conduction band by emitting phonons much more quickly than the radiative transition. Similarly, the holes also relax to the top of the valence band. In the case of crystalline semiconductors without defects or impurities, there is no localized state in the band gap and PL occurs by transition between the bottom of the conduction band and the top of the valence band. In this case the k -selection rule, $k_{\text{photon}} = k_i - k_f$, must be satisfied, where k_{photon} , k_i and k_f denote the wavenumbers of photons, and elec-

trons in the initial and final states, respectively. Since k_{photon} is much smaller than k_i and k_f , we can rewrite the selection rule as $k_i = k_f$. Semiconductors satisfying this condition are called direct-gap semiconductors. Crystalline silicon is one semiconductor in which the direct transition is not allowed by the k -selection rule (indirect-gap semiconductors), but the transition is allowed by either absorption of phonons or their emission. On the other hand, strong PL is observed in a-Si:H. In amorphous semiconductors, the k -selection rule is relaxed and, furthermore, the electrons and holes relax to localized states in the gap before radiative recombination. Thus, PL occurs by transitions between localized states.

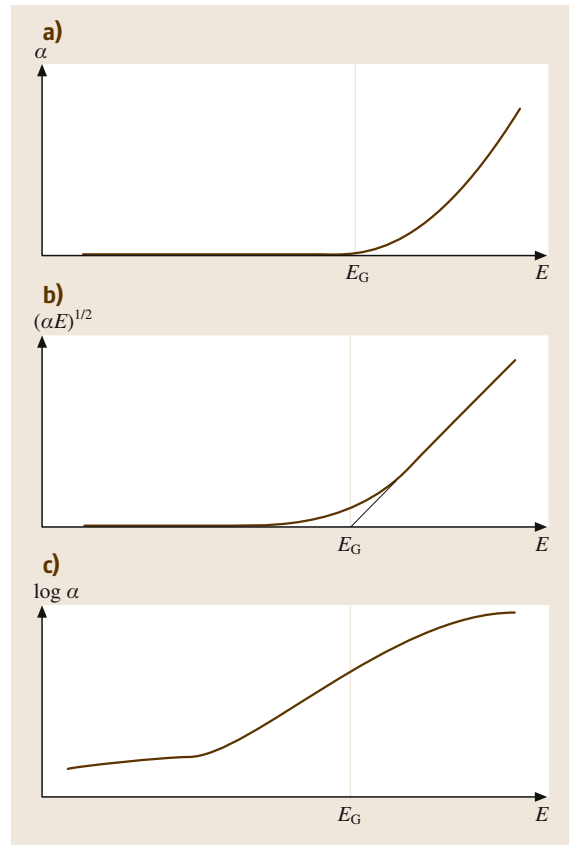


Fig. 25.11a–c Schematic illustration of the absorption spectra of amorphous semiconductors. **(a)** The absorption coefficient α plotted as a function of photon energy E . **(b)** $(\alpha E)^{1/2}$ versus E (Tauc's plot). **(c)** $\log \alpha$ versus E . E_G is the optical gap determined from (b)

The lifetime measurement is important for the identification of the origin of the PL since the PL spectra in most amorphous semiconductors are featureless and do not provide much information. The probability of radiative transition between localized electrons and holes depends exponentially on their separation, R . The lifetime is estimated by $\tau = \tau_0 \exp(2R/R_0)$ where R_0 denotes the radius of the most extended wavefunction [25.36]. When the Coulomb interaction is strong enough, excitons are formed. Exciton formation is also suggested in some amorphous semiconductors. The total spin of excitons, S , is either 0 or 1. Singlet excitons of $S = 0$ have a lifetime in the nanosecond region, while triplet excitons with $S = 1$ have a lifetime much longer than that of singlet excitons, e.g. $\tau = 1$ ms, since the transition is forbidden by the spin selection rule.

PL measurements combined with magnetic resonance are a powerful means to study recombination processes [25.14, 15]. This is called optically detected magnetic resonance (ODMR) measurements. PL from triplet excitons has been suggested from ODMR measurements, as described in Sects. 25.3.2 and 25.3.3.

The photoconductivity measurements provide us with useful information about the processes of carrier transport and recombination. Their results on a-Si:H and chalcogenide glasses are briefly reviewed in Sects. 25.3.2 and 25.3.3, respectively.

Table 25.1 Optical gap of a-Si:H prepared at various deposition temperatures

T_s (°C)	[H] (at.%)	E_g (eV)	References
300	7	1.70	[25.37, 38]
200	18	1.75	[25.37, 38]
120	28	1.9	[25.37, 38]
75	33	2.0	[25.38, 39]

T_s : Deposition temperature, [H]: Hydrogen content

25.3.2 a-Si:H and Related Materials

The band gap of a-Si:H determined from Tauc's plot is 1.7–1.9 eV, depending on the preparation condition, particularly the deposition temperature, T_s , i.e. hydrogen content, [H], [25.37–39, 48] as shown in Table 25.1. The absorption coefficient of visible light for a-Si:H is of the order 10^5 – 10^6 cm⁻¹, which is large compared to that of crystalline silicon. A thickness of 1 μ m is enough to absorb visible light. Thus a-Si:H is suitable for application to solar cells especially when a thin film is desired.

The optical gap, E_g , and the Urbach tail width, E_U , of a-Si:H, a-Ge:H and Si-based alloys are shown in Table 25.2. Some a-Si:H-based alloys such as a-Si_{1-x}N_x:H and a-Si_{1-x}C_x:H have a band gap wider than that of a-Si:H depending on the composition, x . In the preparation of these alloys, x can be varied continuously over a certain range. Thus, the band gap can be varied arbitrarily. It is possible to prepare multilayer films consisting of layers with different band gaps, such as a-Si:H/a-Si_{1-x}N_x:H. Quantum size effects in multilayer films will be described in Sect. 25.6.

It is difficult to obtain the absorption spectra below the band gap from transmittance measurements in a-Si:H, which is normally prepared as a thin film. Such low absorption in a-Si:H has been measured by PDS and CPM.

PL from a-Si:H was observed by Engemann and Fischer [25.49] for the first time. The quantum efficiency has been found to be of the order of unity [25.50], although it decreases with increasing density of dangling bonds [25.51]. PL spectra from a-Si:H films generally consist of two components. The first is observed as a peak at 1.3–1.4 eV with a FWHM of 0.3 eV. The other component, which is called low-energy PL (defect PL), is observed at 0.8–0.9 eV.

The origin of the main peak has not been fully understood. Electrons and holes in the tail states may give

Table 25.2 Band gap energies, Urbach tail widths and dark conductivities at room temperature for a-Si:H, a-Ge:H and related materials

Material	E_g (eV)	E_U (meV)	$\sigma(300\text{ K})$ ($\Omega^{-1}\text{cm}^{-1}$)	References
a-Si:H	1.75	48	10^{-11}	[25.40, 41]
a-Ge:H	1.05	50	10^{-4}	[25.40, 41]
a-Si _{0.7} C _{0.3} :H	2.28	183.4		[25.42]
a-Si _{0.8} C _{0.2} :H	2.2		10^{-15}	[25.43]
a-Si _{0.4} N _{0.6} :H	3.0	≈ 200	$\approx 10^{-8}$	[25.44–46]
a-Si _{0.74} N _{0.26} :H	≈ 2.0	≈ 100	1.4×10^{-8}	[25.44, 45, 47]

rise to this PL. Dunstan and Boulitrop [25.52] have shown that the PL spectra of a-Si:H can be understood without considering electron–phonon coupling. However, the electron–phonon coupling has still been discussed. Morigaki proposed a model on the basis of ODMR measurements [25.14] in which self-trapped holes and tail electrons are the origin of the PL at low temperatures [25.53], while tail holes and tail electrons participate in the PL at high temperatures.

In the case of radiative recombination of localized electron–hole pairs the lifetime depends exponentially on the separation of the electron and the hole, as described in Sect. 25.3.1. When the spatial distribution of electrons and holes is random, a broad lifetime distribution is predicted. This model describes some properties well, e.g., the generation-rate dependence of the lifetime. However, it has been pointed out that we have to consider the PL from specific electronic states such as excitonic PL to understand the lifetime distribution at low generation rate [25.54–56]. PL due to triplet excitons in a-Si:H has been suggested by optically detected magnetic resonance measurements [25.57, 58]. The PL from singlet excitons, which is expected to have lifetime of about 10 ns, has also been reported [25.59–61].

The low-energy PL is emission from deep gap states created by defects. However the origin of low-energy PL is not neutral dangling bonds (T_3^0) of Si since they act as nonradiative centers, as has been suggested from optically detected magnetic resonance [25.62]. Yamaguchi et al. [25.17, 18] proposed a model in which the origin of the low-energy PL is $T_3^+ - N_2^-$ pairs.

The drift mobility of carriers, i.e., electrons and holes, in a-Si:H has been measured by the method of time of flight (TOF), using blocking electrodes [25.63]. In the measurement, carriers are created by an optical pulse near one of the electrodes and run as a sheet-like shape against another electrode. The transient photocurrent i_p

associated with a pulsed optical excitation exhibits a dispersive behavior, as shown in Fig. 25.12 [25.63, 64], which is given by

$$i_p \propto t^{-(1-\alpha)} \quad \text{for } t < t_r \quad (25.3)$$

and

$$i_p \propto t^{-(1+\alpha)} \quad \text{for } t > t_r, \quad (25.4)$$

where t_r designates the transit time, as shown in Fig. 25.12, given by

$$t_r = \frac{L}{\mu_d F}, \quad (25.5)$$

in which L , μ_d and F are the separation of the two electrodes, the drift mobility and the magnitude of the electric field, respectively. In dispersive conduction, the drift mobility depends on the thickness of the samples. This is due to dispersive carrier processes, which are caused by dispersion of the flight time. Band conduction with multiple trapping of carriers also exhibits dispersive conduction, as was observed in a-Si:H and chalcogenide glasses. For undoped a-Si:H, the drift mobility was on the order of $1 \text{ cm}^2/\text{V}$ at room temperature. $\alpha = 0.51$ was obtained as 160 K [25.11]. This transient behavior depends on temperature, i.e., it becomes more dispersive at lower temperatures. α is related to the exponential tail width (normally equal to E_U) $E_c \equiv k_B T_c$ as follows:

$$\alpha = \frac{T}{T_c}. \quad (25.6)$$

From the value of α , the tail width of the conduction band has been obtained to be 25 meV. For hole transport, dispersive behavior has also been obtained and a drift mobility of the order of $10^{-2} \text{ cm}^2/\text{V}$ has been obtained at room temperature.

For the steady-state photoconductivity of undoped a-Si:H, it has been generally accepted that photoconduction occurs through electrons in the conduction band at temperatures above $\approx 60 \text{ K}$ and through hopping of tail electrons at temperatures below $\approx 60 \text{ K}$. The detailed processes of trapping and recombination of carriers involved in photoconduction have been discussed in many literatures (see, e.g., Morigaki [25.2]; Singh and Shimakawa [25.3]). In relation to this issue, spin-dependent photoconductivity measurements are very useful, and have therefore been extensively performed on a-Si:H [25.65–67].

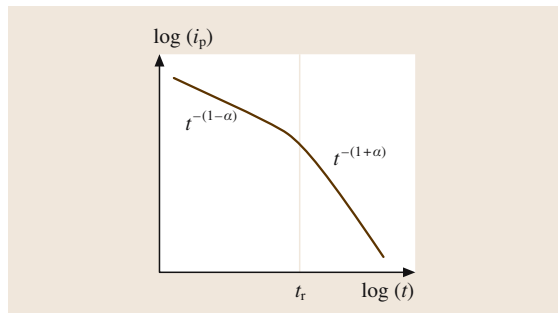


Fig. 25.12 Schematic illustration of the transient photocurrent curve for the dispersive transport

25.3.3 Chalcogenide Glasses

As mentioned in Sect. 25.3.1, the absorption coefficient for the interband transition in many chalcogenide

Table 25.3 Band gap energies, drift mobilities, and dark conductivities at room temperature for chalcogenide glasses [25.2]

Material	E_g [25.68] (eV)	μ_d (cm ² /Vs) [25.1, 69]		σ (300 K) [25.68] ($\Omega^{-1}\text{cm}^{-1}$)
		n-type	p-type	
Se	2.05	$3-6 \times 10^{-3}$	$1-2 \times 10^{-1}$	10^{-16}
As ₂ S ₃	2.32			10^{-17}
As ₂ Se ₃	1.76		10^{-3}	10^{-12}
As ₂ Te ₃	0.83		10^{-2}	10^{-4}
GeSe ₂	3.07			10^{-14}
GeSe ₂	2.18	1.4×10^{-1}	4×10^{-2}	10^{-11}
Sb ₂ Se ₃	0.70			
Cd–In–S [25.70]	2.2	30 *)		9×10^{-1}

*) Hall mobility

glasses is well described by Tauc's relationship (25.1). The optical gap E_g is shown in Table 25.3 for various chalcogenide glasses. The effect of light on optical absorption in chalcogenide glasses is known as photo-darkening. This effect will be briefly reviewed in Sect. 25.5.

PL has been observed in a-Se, arsenic chalcogenide glasses and germanium chalcogenide glasses. PL in chalcogenide glasses is known to have a large Stokes shift. The peak energy in the PL spectra is approximately half of the band gap. This Stokes shift has been attributed to strong electron–phonon coupling [25.71]. The electronic states responsible for the PL have not been well understood. There has been some experimental evidence

for PL from self-trapped excitons. The observation of triplet excitons has been reported in ODMR measurements for chalcogenide glasses such as As₂S₃ and As₂Se₃ [25.72–74].

Dispersive transport was observed in the transient photocurrent for the first time for chalcogenide glasses such as a-As₂Se₃, in which the hole transport dominates over the electron transport [25.63]. The drift mobility of holes ranges between 10^{-3} – 10^{-5} cm²/V, depending on the temperature and electric field. The zero-field hole mobility at room temperature is about 5×10^{-5} cm²/V. The steady-state photoconduction in a-As₂Se₃ has been considered to be governed by charged structural defects such as VAPs.

25.4 Electrical Properties

25.4.1 General Aspects

Electrical Conductivity

Electrical conduction in amorphous semiconductors consists of band conduction and hopping conduction. Band conduction in undoped amorphous semiconductors is characterized by

$$\sigma = \sigma_0 \exp\left(-\frac{E_a}{k_B T}\right), \quad (25.7)$$

where σ and σ_0 are the electrical conductivity and a prefactor, respectively, and E_a , k_B and T are the activation energy, the Boltzmann constant and the temperature, respectively. E_a is given by either $E_c - E_F$ or $E_F - E_V$, depending on whether electrons or holes are considered,

where E_c , E_V and E_F are the mobility edges of the conduction band and the valence band, and the Fermi energy, respectively. Hopping conduction in amorphous semiconductors consists of nearest-neighbor hopping and variable-range hopping [25.1]. Nearest-neighbor hopping is well known in crystalline semiconductors, in which electrons (holes) hop to nearest-neighbor sites by emitting or absorbing phonons. Variable-range hopping is particularly associated with tail states, in which electrons (holes) in tail states hop to the most probable sites. This type of hopping conductivity σ_p is characterized by the following temperature variation:

$$\sigma_p = \sigma_{p0} \exp\left(-\frac{B}{T^{1/4}}\right). \quad (25.8)$$

The Hall Effect

The Hall effect is used for the determination of carrier density and its sign in crystalline semiconductors. For amorphous semiconductors, however, the sign of the Hall coefficient does not always coincide with that of the carriers [25.76]. Such a sign anomaly has been observed in a-Si:H, as will be shown in Sect. 25.4.2.

Thermoelectric Power

The thermoelectric power S associated with band conduction of electrons is given by

$$S = -\frac{k_B}{e} \left(\frac{E_c - E_F}{k_B T} + A \right), \quad (25.9)$$

where A is a quantity depending on the energy dependence of the relaxation times associated with electrical conduction. For hole conduction, S is given by (25.9) except that $E_c - E_F$ is replaced by $E_F - E_V$ and that the sign of S is positive. The sign of S coincides with that of the carriers. The thermoelectric power S (25.9) is related to the electrical conductivity σ (25.7) as follows [25.77]:

$$\ln \sigma + \left| \frac{e}{k_B} S \right| = \ln \sigma_0 + A \equiv Q. \quad (25.10)$$

The temperature-dependent quantity Q is defined as above. The activation energies of σ and S , E_σ and E_S , are equal to each other. Then, Q is defined by Q_0 as follows:

$$Q = \ln \sigma_0 + A \equiv Q_0. \quad (25.11)$$

However, if E_σ is not equal to E_S , Q is generally expressed by

$$Q = Q_0 - \frac{E_Q}{k_B T} \quad (25.12)$$

$$E_Q = E_\sigma - E_S. \quad (25.13)$$

It has been generally observed that E_σ is greater than E_S , i.e., $E_\sigma > E_S$. This has been accounted for in terms of long-range fluctuation of the band edge. The electrical conduction is due to those carriers that are thermally excited into the valley of the band-edge fluctuation and cross over its barrier, while the thermoelectric power, i.e., transport of phonon energy by carriers, is governed by carriers in the valleys. Thus, a relationship of $E_\sigma > E_S$ can be accounted for in terms of long-range fluctuation of the band edge [25.77].

Drift Mobility

The drift mobility of carriers is measured by the TOF method, as mentioned in Sect. 25.3.2. The drift mobility μ_d is estimated from (25.5).

In the following, the electrical properties of a-Si:H and related materials and chalcogenide glasses are described.

25.4.2 a-Si:H and Related Materials

A typical example of band conduction and variable-range hopping conduction is shown in Fig. 25.13, which is the temperature dependence of the electrical conductivity for a-Si_xAu_{1-x} films prepared by vapor evaporation [25.75]. Band conduction occurs at high temperatures and variable-range hopping conduction at low temperatures. The $T^{-1/4}$ law of temperature variation of electrical conductivity is clearly seen in Fig. 25.14. The values of the dark conductivities of a-Si:H, a-Ge:H and related materials at room temperature are shown in Table 25.2. Figure 25.15 shows the temperature dependencies of S for a-Ge:H, in which two slopes of the curve of S versus T^{-1} are seen with 0.43 eV in the high-temperature range and 0.17 eV in the low-temperature range [25.78]. This crossover of two slopes is accounted for in terms of two different mechanisms of electrical conduction, i.e., band conduction at high temperatures and hopping conduction in the band-tail states at low temperatures [25.79]. In amorphous semiconductors, doping control is generally difficult, because the constituent

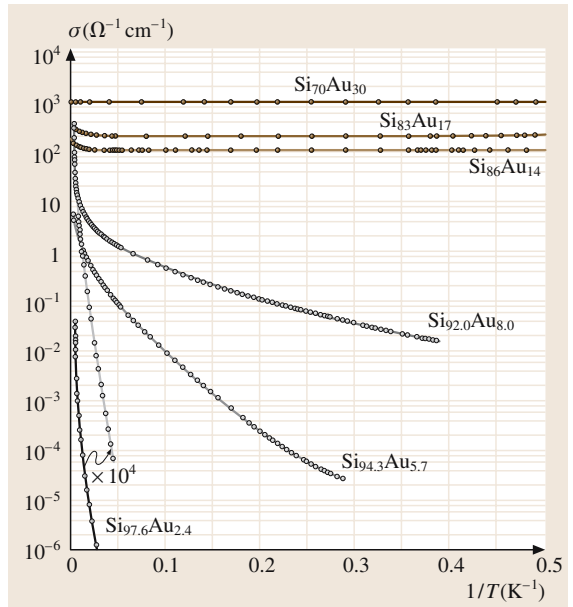


Fig. 25.13 Temperature dependence of the electrical conductivity for an a-Si_xAu_{1-x} film. (After [25.75])

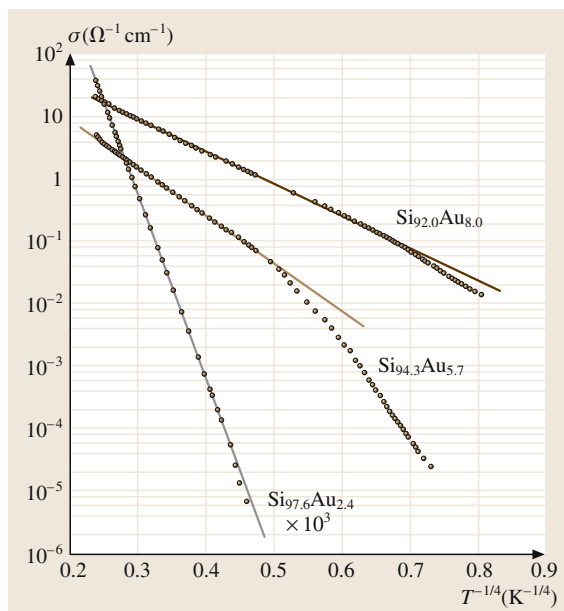


Fig. 25.14 Electrical conductivity versus $T^{-1/4}$ for $a\text{-Si}_x\text{Au}_{1-x}$ films. (After [25.75])

atoms obey the so called $8 - N$ rule to satisfy the coordination of covalent bonds with their neighboring atoms. In $a\text{-Si:H}$, however, n-type doping and p-type doping were performed by using phosphorus (group V element) and boron (group III element), respectively [25.80].

The Hall coefficient of $a\text{-Si:H}$ has been measured [25.76]. In crystalline semiconductors, when the carriers are electrons, its sign is negative, and when the carriers are holes, its sign is positive. The sign of the Hall coefficient is generally negative irrespective of that of the carriers in amorphous semiconductors. This has been accounted for in terms of the random phase model [25.82]. However, for the Hall coefficient of $a\text{-Si:H}$, this is not the case, i.e., when the carriers are electrons, the sign is positive, and when the carriers are holes, the sign is negative. Such an anomalous Hall coefficient in $a\text{-Si:H}$ is called double reversal of the Hall coefficient and has been accounted for, taking into account antibonding orbitals for electrons and bonding orbitals for holes [25.83].

25.5 Light-Induced Phenomena

Light-induced phenomena in amorphous semiconductors were first observed for chalcogenide glasses

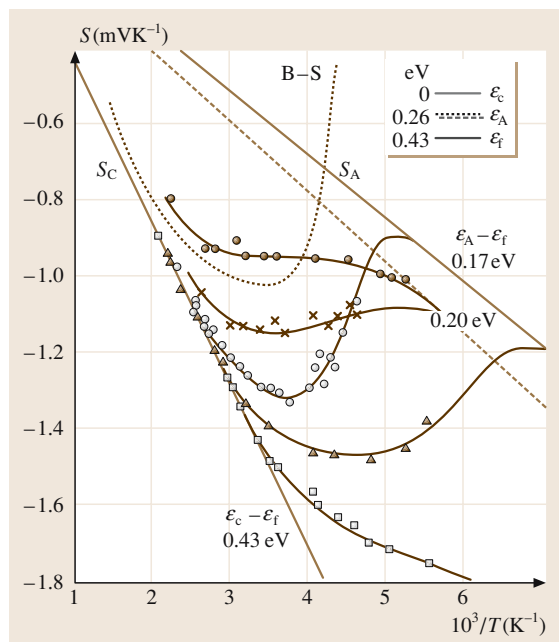


Fig. 25.15 Thermoelectric power versus T^{-1} for various $a\text{-Ge:H}$ films. The solid lines are the calculated results. The insert shows the model on which the calculations are based. Carriers in states just above E_A move in extended states, those at E_A move by hopping. The dotted curve marked B-S was obtained by Beyer and Stuke [25.81] for a slowly evaporated sample annealed at 310 °C [25.78]

25.4.3 Chalcogenide Glasses

The values of the dark conductivities of chalcogenide glasses at room temperature are shown in Table 25.3 along with those of μ_d at room temperature. In chalcogenide glasses, the Fermi level is pinned near the midgap and this has been considered to be due to VAP defects with negative correlation energy. Thus, doping seems difficult, except for a few cases, e.g., the incorporation of Bi into $a\text{-Ge:S}$ and $a\text{-Ge:Se}$ increases the direct current (DC) conductivity by 6–7 orders of magnitude [25.84, 85]. This doping effect is due to chemical modification instead of conventional doping. Furthermore, Cd–In–S chalcogenide glasses exhibit $\sigma \approx 1 \times 10^{-2} \Omega^{-1}\text{cm}^{-1}$ at room temperature [25.70].

as photo-darkening, PL fatigue and photostructural changes associated with illumination whose energy

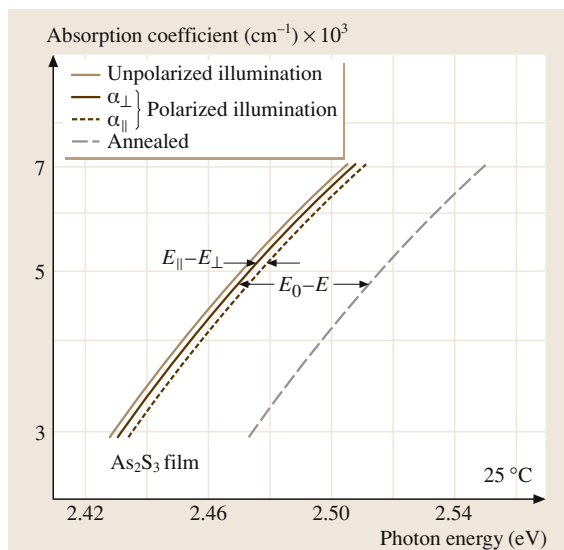


Fig. 25.16 Optical absorption edge of As_2S_3 films annealed or illuminated with unpolarized or linearly polarized light (α_\perp : solid line and α_\parallel : dashed line). α_\parallel and α_\perp refer to polarization along the plane of illumination and perpendicular to it, respectively. (After [25.87])

equals to or exceeds the band-gap energy (see, e.g., Shimakawa et al. [25.86]; Singh and Shimakawa, [25.3]). Photo-darkening is an effect in which the absorption edge shifts towards lower photon energy under illumination, as shown in Fig. 25.16 [25.87]. PL fatigue is the decrease of the PL intensity associated with illumination. Photostructural changes are observed as changes of the volume (generally expansion) of the sample with illumination. PL fatigue may be due to light-induced creation of charged structural defects acting as nonradiative recombination centers. For the photo-darkening and the photostructural change, it has been discussed whether they are independent of each other or not. Thus, the origins of these phenomena are still unclear.

Light-induced phenomena in a-Si:H were observed first in the dark conductivity and photoconductivity, i.e., drops in their values after prolonged illumination of band-gap light (the so-called Staebler–Wronski effect [25.88]). Subsequently, such phenomena have been observed in PL, optical absorption and ODMR etc. [25.2]. This effect also gives rise to degradation of the performance of amorphous-silicon solar cells, so that this has received great attention from the viewpoint of application. After prolonged illumination, dangling bonds were found to be created [25.89,90], so that light-induced creation of dangling bonds has been considered

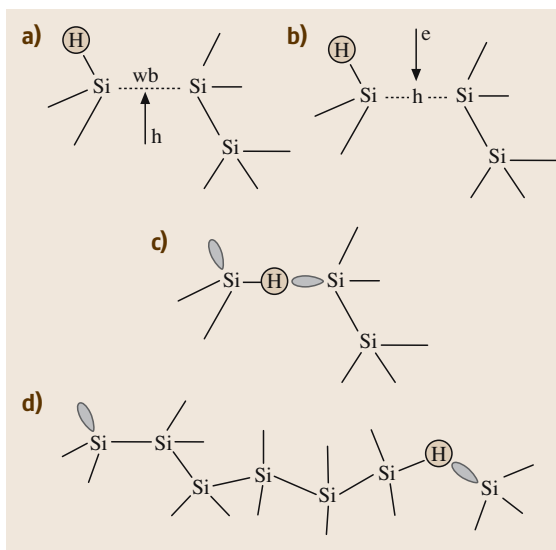


Fig. 25.17 (a) Self-trapping of a hole by a weak Si–Si bond adjacent to a Si–H bond. (b) Nonradiative recombination of a self-trapped hole with an electron. (c) A Si–H bond is switched toward the weak Si–Si bond and a dangling bond is left behind. (d) Formation of two separate dangling bonds after hydrogen movements and repeating of processes shown in (a)–(c)

as the origin of these phenomena. Several models have been proposed for mechanisms for light-induced creation of dangling bonds [25.53, 91–96], but this issue is still controversial (see, e.g., Morigaki [25.2]; Singh and Shimakawa [25.3]). Very recently, the creation of a number of dangling bonds such as $1 \times 10^{19} \text{ cm}^{-3}$ after pulsed optical excitation has been observed in high-quality a-Si:H films [25.97, 98]. This result has been accounted for in terms of the authors' model [25.99], based on a combination of the following processes occurring during illumination. Self-trapping of holes in weak Si–Si bonds adjacent to Si–H bonds triggers these weak bonds to break using the phonon energy associated with nonradiative recombination of electrons with those self-trapped holes (Fig. 25.17a and b). After Si–H bond switching and hydrogen movement, two types of dangling bonds are created, i.e., a normal dangling bond and a dangling bond with hydrogen at a nearby site, i.e. so-called hydrogen-related dangling bonds (Fig. 25.17c and d). In the latter, hydrogen is dissociated from the Si–H bond as a result of nonradiative recombination at this dangling bond site. Dissociated hydrogen can terminate two types of dangling bonds (Fig. 25.18c and d) or can be in-

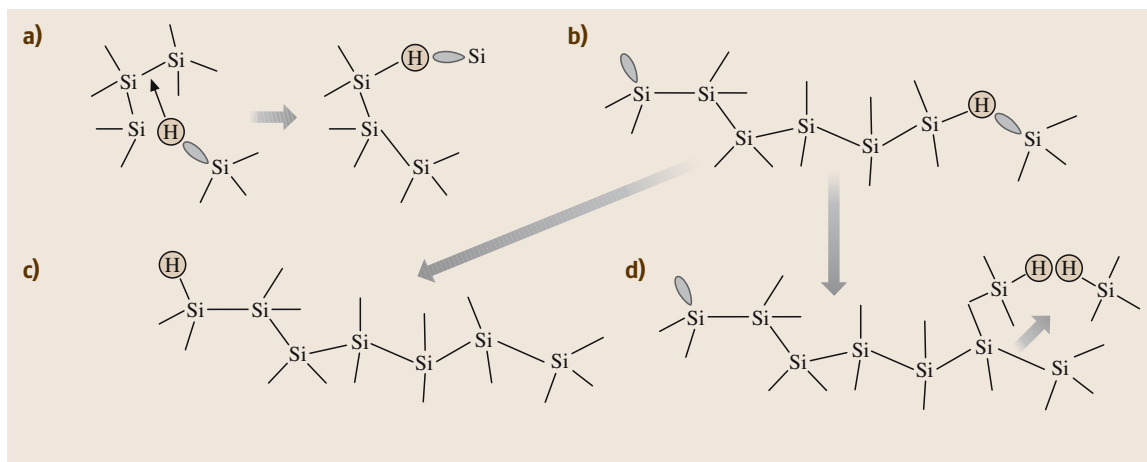


Fig. 25.18 (a) Dissociation of hydrogen from a hydrogen-related dangling bond and insertion of hydrogen into a nearby weak Si—Si bond (formation of a new hydrogen-related dangling bond). (b) Two separate dangling bonds, i. e. a normal dangling bond and a hydrogen-related dangling bond. (c) Dissociation of hydrogen from a hydrogen-related dangling bond and termination of a normal dangling bond by hydrogen. (d) Termination of a hydrogen-related dangling bond by hydrogen

serted into a nearby weak Si—Si bond (Fig. 25.18a). The dependencies of the density of light-induced dangling bonds on illumination time and generation rate have been calculated using the rate equations governing these processes. The results have been compared with experimental results obtained under continuous

illumination and pulsed illumination with good agreement. Light-induced structural changes, i. e., volume change [25.100] and local structural changes around a Si—H bond [25.101] have also been observed in a-Si:H and the origins for these changes have been discussed [25.102–104].

25.6 Nanosized Amorphous Structure

Recent technologies have enabled us to control the properties of semiconductors by introducing artificial structures of nanometer size such as quantum wells, quantum wires and quantum dots. The quantum well is formed by preparing multilayers consisting of two semiconductors. Figure 25.19a illustrates the conduction and valence band edges in such multilayers consisting of two semiconducting materials with different band gaps, plotted as functions of the distance from the substrate, z . The curves of the conduction and valence bands in Fig. 25.19a are considered to be the potentials for the electrons and holes, respectively. The motion of the electrons and holes is assumed to be described by effective mass equations which are similar to the Schrödinger equation. When the barriers of the potentials are high enough to prevent the carriers from moving to the adjacent well, the quantum levels (for motion along the z -axis) are as illustrated in Fig. 25.19a. The band gap

of the multilayer, E_g , is equal to the separation of the quantum levels of the lowest energy in the conduction and valence bands, as shown in Fig. 25.19a. E_g increases with decreasing thickness of the well layer, L_w ; this is well known as the quantum-size effect. When the height of the barrier is infinite, $\Delta E = E_g - E_{gw}$ is proportional to L_w^{-2} , where E_{gw} denotes the band gap of the material of the well layer.

Multilayers of various amorphous semiconductors have been prepared. Observation of the quantum size effect has been reported in some amorphous semiconducting multilayers. For example, the shift of the optical gap of an a-Si:H/a-Si_{1-x}N_x:H multilayer from that of a-Si:H, ΔE , is proportional to L_w^{-2} , where L_w is the thickness of the a-Si:H layers (e.g. Morigaki [25.2]; see Yamaguchi and Morigaki [25.105]; and references therein). The result, $\Delta E \propto L_w^{-2}$, is consistent with quantum size effect in a square-well potential.

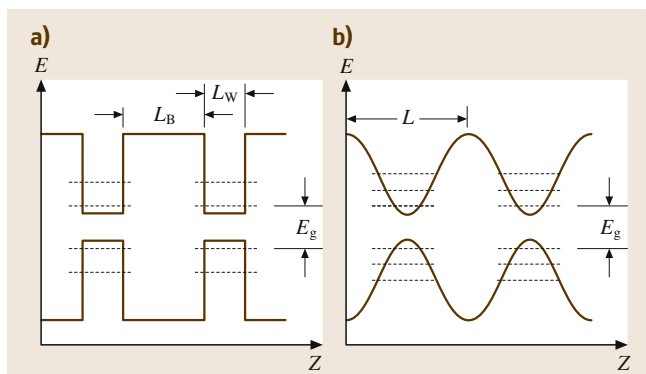


Fig. 25.19a,b The edges of the conduction and valence bands in (a) multilayer films and (b) band-edge modulated films, as functions of z , where z denotes distance from the substrates. The dashed lines indicate quantum levels. The band gap E_g , which increases due to the quantum size effect, is also shown

In the case of amorphous semiconductors, it is easy to introduce a potential well of arbitrary shape, because the optical gap of the amorphous semiconducting alloy, which depends on the composition, can be varied continuously within a certain range. The authors have prepared a new type of a-Si:H-based film called band-edge modulated (BM) a-Si_{1-x}N_x:H, in which the potentials for electrons and holes are sinusoidal functions of z [25.106,107]. The band-edges for the BM films are illustrated in Fig. 25.19b. In this case, the quantum

size effect is also expected, similarly to the case of multilayer films. However, the quantum levels and the size dependencies of the band gap are different from those in the case of multilayer films. In the case of BM, the potential is approximately parabolic at the bottom of the well. In this case the quantum levels have the same separation as that illustrated in Fig. 25.19b. We expect that the shift of the optical gap in BM films, ΔE , will be proportional to L^{-1} , where L denotes the modulation period. The band gap of BM films observed in experiments are in agreement with that expected from the above consideration [25.106]. Parabolic potentials in crystalline semiconductors have also been introduced (see Gossard et al. [25.108]). However it is difficult to obtain such structures that are small enough to observe significant quantum effects in crystalline semiconductors.

Recently, preparation of amorphous silicon quantum dots in silicon nitride has been reported [25.109]. Theoretical calculations for silicon nanostructures [25.110] have predicted that the oscillator strength in low symmetry is larger than that in high symmetry. A large quantum efficiency for PL is expected in the case of amorphous quantum dots. Thus, amorphous quantum wires and dots have potential applications in light-emitting devices with high efficiency.

Chalcogenide glasses multilayers have been prepared [25.111–113], whose optical properties such as optical absorption and PL, and electrical properties have been measured.

References

- 25.1 N. F. Mott, E. A. Davis: *Electronic Processes in Non-crystalline Materials*, 2nd edn. (Clarendon, Oxford 1979)
- 25.2 K. Morigaki: *Physics of Amorphous Semiconductors* (World Scientific, Singapore; Imperial College Press, London 1999)
- 25.3 J. Singh, K. Shimakawa: *Advances in Amorphous Semiconductors* (Taylor Francis, London 2003)
- 25.4 S. Kugler, P. R. Surján, G. Náray-Szabó: *Phys. Rev. B* **37**, 9069 (1988)
- 25.5 L. Ley, J. Reichardt, R. L. Johnson: *Phys. Rev. Lett.* **49**, 1664 (1982)
- 25.6 L. Brey, C. Tejedor, J. A. Verges: *Phys. Rev. Lett.* **52**, 1840 (1984)
- 25.7 L. Ley, S. Kowalczyk, R. Pollak, D. A. Shirley: *Phys. Rev. Lett.* **29**, 1088 (1972)
- 25.8 M. Kastner, D. Adler, H. Fritzsche: *Phys. Rev. Lett.* **37**, 1504 (1976)
- 25.9 K. Morigaki, M. Yamaguchi, I. Hirabayashi, R. Hayashi: *Disordered Semiconductors*, ed. by M. A. Kastner, G. A. Thomas, S. R. Ovshinsky (Plenum, New York 1987) p. 415
- 25.10 G. D. Cody: In: *Semiconductors and Semimetals*, Vol. 21, Part B ed. by J. I. Pankove (Academic, Orlando 1984) p. 1
- 25.11 T. Tiedje: In: *Semiconductors and Semimetals*, Vol. 21, Part C ed. by J. I. Pankove (Academic, Orlando 1984) p. 207
- 25.12 K. Winer, I. Hirabayashi, L. Ley: *Phys. Rev. B* **38**, 7680 (1988)
- 25.13 K. Winer, I. Hirabayashi, L. Ley: *Phys. Rev. Lett.* **60**, 2697 (1988)
- 25.14 K. Morigaki: In: *Semiconductors and Semimetals*, Vol. 21, Part C ed. by J. I. Pankove (Academic, Orlando 1984) p. 155
- 25.15 K. Morigaki, M. Kondo: *Solid State Phenomena* **44–46**, 731 (1995)
- 25.16 K. Morigaki, H. Hikita, M. Kondo: *J. Non-Cryst. Solids* **190**, 38 (1995)

- 25.17 M. Yamaguchi, K. Morigaki, S. Nitta: J. Phys. Soc. Jpn. **58**, 3828 (1989)
- 25.18 M. Yamaguchi, K. Morigaki, S. Nitta: J. Phys. Soc. Jpn. **60**, 1769 (1991)
- 25.19 S. C. Moss, J. F. Graczyk: *Proc. 10th Int. Conf. on Physics of Semiconductors*, ed. by J. C. Hensel, F. Stern (US AEC Div. Tech. Inform., Springfield 1970) p. 658
- 25.20 S. Kugler, G. Molnár, G. Pető, E. Zsoldos, L. Rosta, A. Menelle, R. Bellissent: Phys. Rev. B **40**, 8030 (1989)
- 25.21 D. E. Polk: J. Non-Cryst. Solids **5**, 365 (1971)
- 25.22 F. Wooten, K. Winer, D. Weaire: Phys. Rev. Lett. **54**, 1392 (1985)
- 25.23 R. Car, M. Parrinello: Phys. Rev. Lett. **60**, 204 (1988)
- 25.24 S. Kugler, L. Pusztai, L. Rosta, P. Chieux, R. Bellissent: Phys. Rev. B **48**, 7685 (1993)
- 25.25 L. Pusztai: J. Non-Cryst. Solids **227-230**, 88 (1998)
- 25.26 S. Kugler, K. Kohary, K. Kádas, L. Pusztai: Solid State Commun. **127**, 305 (2003)
- 25.27 J. A. Reimer: J. Phys. (Paris) **42**, C4-715 (1981)
- 25.28 K. K. Gleason, M. A. Petrich, J. A. Reimer: Phys. Rev. B **36**, 3259 (1987)
- 25.29 G. Lucovsky, R. J. Nemanich, J. C. Knights: Phys. Rev. B **19**, 2064 (1979)
- 25.30 R. Bellissent, A. Menelle, W. S. Howells, A. C. Wright, T. M. Brunier, R. N. Sinclair, F. Jansen: Physica B **156**, 157, 217 (1989)
- 25.31 A. Menelle: Thèse Doctorat (Université Pierre et Marie Curie, Paris 1987)
- 25.32 T. Uchino: Kotai Butsuri (Solid State Physics) **37**, 965 (2002)
- 25.33 J. H. Zhou, K. Ikuta, T. Yasuda, T. Umeda, S. Yamasaki, K. Tanaka: J. Non-Cryst. Solids **227-230**, 857 (1998)
- 25.34 G. Lucovsky, F. L. Galeener, R. H. Geils, R. C. Keezer: In: *Proc. Int. Conf. on Amorphous and Liquid Semiconductors*, ed. by W. E. Spear (University of Edinburgh, Edinburgh 1977) p. 127
- 25.35 J. Tauc: In: *Amorphous and Liquid Semiconductors*, ed. by J. Tauc (Plenum, New York 1974) p. 159
- 25.36 C. Tsang, R. A. Street: Phys. Rev. B **19**, 3027 (1979)
- 25.37 K. Morigaki, Y. Sano, I. Hirabayashi: J. Phys. Soc. Jpn. **51**, 147 (1982)
- 25.38 K. Morigaki, Y. Sano, I. Hirabayashi: *Amorphous Semiconductor Technologies and Devices-1983*, ed. by Y. Hamakawa (Ohomsha, Tokyo 1983) Chap. 3.2
- 25.39 I. Hirabayashi, K. Morigaki, M. Yoshida: Sol. Ener. Mat. **8**, 153 (1982)
- 25.40 G. H. Bauer: Solid State Phenomena **44-46**, 365 (1995)
- 25.41 W. Paul: In: *Amorphous Silicon and Related Materials*, ed. by H. Fritzsche (World Scientific, Singapore 1989) p. 63
- 25.42 Y. Tawada: In: *Amorphous Semiconductors—Technologies and Devices*, ed. by Y. Hamakawa (Ohmsha and North Holland, Tokyo and Amsterdam 1983) Chap. 4.2
- 25.43 F. Demichelis, C. F. Pirri: Solid State Phenomena **44-46**, 385 (1995)
- 25.44 K. Maeda, I. Umezu: J. Appl. Phys. **70**, 2745 (1991)
- 25.45 M. Yamaguchi, K. Morigaki unpublished
- 25.46 B. Dunnett, D. I. Jones, A. D. Stewart: Philos. Mag. B **53**, 159 (1986)
- 25.47 M. Hirose: Jpn. J. Appl. Phys. **21**(suppl. 21-1), 297 (1981)
- 25.48 M. Yamaguchi, K. Morigaki: Philos. Mag. B **79**, 387 (1999)
- 25.49 D. Engemann, R. Fischer: In: *Amorphous and Liquid Semiconductors*, ed. by J. Stuke, W. Brenig (Taylor & Francis, London 1974) p. 947
- 25.50 D. Engemann, R. Fischer: In: *Proceedings of the 12th International Conference on the Physics of Semiconductors*, ed. by M. H. Pilkuhn (B. G. Teubner, Stuttgart 1974) p. 1042
- 25.51 R. A. Street, J. C. Knights, D. K. Biegelsen: Phys. Rev. B **18**, 1880 (1978)
- 25.52 D. J. Dunstan, F. Boulitrop: Phys. Rev. B **30**, 5945 (1984)
- 25.53 K. Morigaki: J. Non-Cryst. Solids **141**, 166 (1992)
- 25.54 R. Stachowitz, M. Schubert, W. Fuhs: J. Non-Cryst. Solids **227-230**, 190 (1998)
- 25.55 C. Ogihara: J. Non-Cryst. Solids **227-230**, 517 (1998)
- 25.56 T. Aoki, T. Shimizu, S. Komodoori, S. Kobayashi, K. Shimakawa: J. Non-Cryst. Solids **338-340**, 456 (2004)
- 25.57 M. Yoshida, M. Yamaguchi, K. Morigaki: J. Non-Cryst. Solids **114**, 319 (1989)
- 25.58 M. Yoshida, K. Morigaki: J. Phys. Soc. Jpn. **58**, 3371 (1989)
- 25.59 B. A. Wilson, P. Hu, T. M. Jedju, J. P. Harbison: Phys. Rev. B **28**, 5901 (1983)
- 25.60 C. Ogihara, H. Takemura, H. Yoshida, K. Morigaki: J. Non-Cryst. Solids **266-269**, 574 (2000)
- 25.61 H. Takemura, C. Ogihara, K. Morigaki: J. Phys. Soc. Jpn. **71**, 625 (2002)
- 25.62 M. Yoshida, K. Morigaki: J. Non-Cryst. Solids **59 & 60**, 357 (1983)
- 25.63 G. Pfister, H. Scher: Adv. Phys. **27**, 747 (1978)
- 25.64 H. Scher, E. W. Montroll: Phys. Rev. B **12**, 2455 (1975)
- 25.65 I. Solomon: In: *Amorphous Semiconductors*, ed. by M. H. Brodsky (Springer, Berlin Heidelberg New York 1979) p. 189
- 25.66 K. Lips, C. Lerner, W. Fuhs: J. Non-Cryst. Solids **198-200**, 267 (1996)
- 25.67 M. Stutzmann, M. S. Brandt, M. W. Bayerl: J. Non-Cryst. Solids **266-269**, 1 (2000)
- 25.68 S. R. Elliott: In: *Material Science and Technology*, Vol. 9, ed. by R. W. Cahn et al. (VCH, Weinheim 1991) p. 376
- 25.69 A. Feltz: *Amorphous Inorganic Materials and Glasses* (VCH, Weinheim 1993)
- 25.70 H. Hosono, H. Maeda, Y. Kameshima, H. Kawazoe: J. Non-Cryst. Solids **227-230**, 804 (1998)
- 25.71 R. A. Street: Adv. Phys. **25**, 397 (1976)

- 25.72 B. C. Cavenett: J. Non-Cryst. Solids **59 & 60**, 125 (1983)
- 25.73 J. Ristein, P. C. Taylor, W. D. Ohlsen, G. Weiser: Phys. Rev. B **42**, 11845 (1990)
- 25.74 D. Mao, W. D. Ohlsen, P. C. Taylor: Phys. Rev. B **48**, 4428 (1993)
- 25.75 N. Kishimoto, K. Morigaki: J. Phys. Soc. Jpn. **46**, 846 (1979)
- 25.76 P. G. LeComber, D. I. Jones, W. E. Spear: Philos. Mag. **35**, 1173 (1977)
- 25.77 H. Overhof, W. Beyer: Philos. Mag. B **44**, 317 (1983)
- 25.78 D. I. Jones, W. E. Spear, P. G. LeComber: J. Non-Cryst. Solids **20**, 259 (1976)
- 25.79 N. F. Mott: J. Phys. C **13**, 5433 (1980)
- 25.80 W. E. Spear, P. G. LeComber: Philos. Mag. **33**, 935 (1976)
- 25.81 W. Beyer, J. Stuke: In: *Proc. Int. Conf. on Amorphous and Liquid Semiconductors, 1973*, ed. by J. Stuke (Taylor & Francis, London 1974) p. 251
- 25.82 L. Friedman: J. Non-Cryst. Solids **6**, 329 (1971)
- 25.83 D. Emin: Philos. Mag. **35**, 1189 (1977)
- 25.84 N. Tohge, T. Minami, Y. Yamamoto, M. Tanaka: J. Appl. Phys. **51**, 1048 (1980)
- 25.85 L. Tichy, H. Ticha, A. Triska, P. Nagels: Solid State Commun. **53**, 399 (1985)
- 25.86 K. Shimakawa, A. Kolobov, S. R. Elliott: Adv. Phys. **44**, 475 (1995)
- 25.87 K. Kimura, K. Murayama, T. Ninomiya: J. Non-Cryst. Solids **77, 78**, 1203 (1985)
- 25.88 D. L. Staebler, C. R. Wronski: Appl. Phys. Lett. **31**, 292 (1977)
- 25.89 I. Hirabayashi, K. Morigaki, S. Nitta: Jpn. J. Appl. Phys. **19**, L357 (1980)
- 25.90 H. Dersch, J. Stuke, J. Beichler: Appl. Phys. Lett. **38**, 456 (1981)
- 25.91 M. Stutzmann, W. B. Jackson, C. C. Tsai: Phys. Rev. B **32**, 23 (1985)
- 25.92 C. Godet, P. Roca i Cabarrocas: J. Appl. Phys. **80**, 97 (1996)
- 25.93 H. M. Branz: Phys. Rev. B **59**, 5498 (1999)
- 25.94 K. Morigaki, H. Hikita: Solid State Commun. **114**, 69 (2000)
- 25.95 K. Morigaki, H. Hikita: J. Non-Cryst. Solids **266-269**, 410 (2000)
- 25.96 K. Morigaki, H. Hikita: *Proc. Int. Conf. on Physics of Semiconductors*, ed. by T. Ando N. Miura (Springer, Berlin Heidelberg New York 2000) p. 1485
- 25.97 C. Ogihara, H. Takemura, T. Yoshimura, K. Morigaki: J. Non-Cryst. Solids **299-302**, 637 (2002)
- 25.98 K. Morigaki, H. Hikita, H. Takemura, T. Yoshimura, C. Ogihara: Philos. Mag. Lett. **83**, 341 (2003)
- 25.99 K. Morigaki, H. Hikita: J. Non-Cryst. Solids **299-302**, 455 (2002)
- 25.100 T. Gotoh, S. Nonomura, M. Nishio, S. Nitta, M. Kondo, A. Matsuda: Appl. Phys. Lett. **72**, 2978 (1998)
- 25.101 Y. Zhao, D. Zhang, G. Kong, G. Pan, X. Liao: Phys. Rev. Lett. **74**, 558 (1995)
- 25.102 H. Fritzsche: Solid State Commun. **94**, 953 (1995)
- 25.103 R. Biswas, Y. P. Li: Phys. Rev. Lett. **82**, 2512 (1999)
- 25.104 K. Morigaki: Res. Bull. Hiroshima Inst. Tech. **35**, 47 (2001)
- 25.105 M. Yamaguchi, K. Morigaki: Phys. Rev. B **55**, 2368 (1997)
- 25.106 C. Ogihara, H. Ohta, M. Yamaguchi, K. Morigaki: Philos. Mag. B **62**, 261 (1990)
- 25.107 M. Yamaguchi, C. Ogihara, K. Morigaki: Mat. Sci. Eng. B **97**, 135 (2003)
- 25.108 A. C. Gossard, M. Sundaram, P. F. Hopkins: In: *Semiconductors and Semimetals*, Vol. 40, ed. by A. C. Gossard (Academic, Boston; Tokio 1994) Chap. 2
- 25.109 N.-M. Park, C.-J. Choi, T.-Y. Seong, S.-J. Park: Phys. Rev. Lett. **86**, 1355 (2001)
- 25.110 J. Koga, K. Nishio, T. Yamaguchi, F. Yonezawa: J. Phys. Soc. Jpn. **70**, 3143 (2001)
- 25.111 R. Ionov, D. Nesheva, D. Arsova: J. Non-Cryst. Solids **137&138**, 1151 (1991)
- 25.112 R. Ionov: Ph. D. Thesis. Ph.D. Thesis (Technical Univ., Sofia 1993)
- 25.113 H. Hamanaka, S. Konagai, K. Murayama, M. Yamaguchi, K. Morigaki: J. Non-Cryst. Solids **198-200**, 808 (1996)

26. Amorphous and Microcrystalline Silicon

Processes used to grow hydrogenated amorphous silicon (a-Si:H) and microcrystalline silicon ($\mu\text{c-Si:H}$) from SiH_4 and H_2/SiH_4 glow discharge plasmas are reviewed. Differences and similarities between growth reactions of a-Si:H and $\mu\text{c-Si:H}$ in a plasma and on a film-growing surface are discussed, and the process of nucleus formation followed by epitaxial-like crystal growth is explained as being unique to $\mu\text{c-Si:H}$. The application of a reaction used to determine the dangling-bond defect density in the resulting a-Si:H and $\mu\text{c-Si:H}$ films is emphasized, since it can provide clues about how to improve the optoelectronic properties of those materials for device applications, especially thin-film silicon-based solar cells. Material issues related to the realization of low-cost and high-efficiency solar cells are described, and finally recent progress in this area is reviewed.

26.1	Reactions in SiH_4 and SiH_4/H_2 Plasmas ..	581
26.2	Film Growth on a Surface	583
26.2.1	Growth of a-Si:H	583
26.2.2	Growth of $\mu\text{c-Si:H}$	584
26.3	Defect Density Determination for a-Si:H and $\mu\text{c-Si:H}$	589
26.3.1	Dangling Bond Defects	589
26.3.2	Dangling Bond Defect Density in $\mu\text{c-Si:H}$	590
26.4	Device Applications	590
26.5	Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells ..	591
26.5.1	Controlling Photoinduced Degradation in a-Si:H	591
26.5.2	High Growth Rates of Device-Grade $\mu\text{c-Si:H}$	592
26.6	Summary	594
	References	594

Hydrogenated amorphous silicon (a-Si:H) and microcrystalline silicon ($\mu\text{c-Si:H}$) are recognized as being useful materials for constructing devices related to optoelectronics, such as solar cells, thin-film transistors, etc. [26.1, 2]. Several methods have been proposed for the preparation of device-grade a-Si:H and $\mu\text{c-Si:H}$. These include: reactive sputtering of a crystalline silicon target with $\text{Ar}+\text{H}_2$ plasma [26.3]; mercury-sensitized photochemical vapor deposition (CVD) utilizing a decomposition reaction of silane (SiH_4) molecules with photoexcited Hg (Hg^*) [26.4]; a direct photo CVD method where high-energy photons from a Xe-resonance lamp or a low-pressure Hg lamp are used for the direct excitation of SiH_4 molecules to excited electronic states [26.5, 6]; a hot-wire CVD method for

decomposing SiH_4 by means of catalytic reactions on a heated metal surface [26.7]; and a plasma-enhanced CVD method (PECVD). The PECVD method is the most widely used of these due to its ability to consistently prepare uniform, high-quality materials on a large-area substrate.

In this chapter, the PECVD method is highlighted, details regarding the processes used to grow a-Si:H and $\mu\text{c-Si:H}$ from reactive plasmas are explained, and the determination reaction (which is used to obtain the dangling-bond defect density in the films: one of the most important structural properties that influences device performance) is interpreted in order to obtain clues about how to control the optoelectronic properties of those materials for device applications.

26.1 Reactions in SiH_4 and SiH_4/H_2 Plasmas

The initial event required for the growth of a-Si:H and $\mu\text{c-Si:H}$ is the decomposition of the source gas material in SiH_4 or SiH_4/H_2 glow discharge plasma. Figure 26.1 shows a schematic of the dissociation pathway of SiH_4

and H_2 , during which the molecules are excited to higher electronic states due to inelastic collisions with high-energy electrons in the plasma [26.8]. As the electrons in the plasma usually have a wide variety of energies,

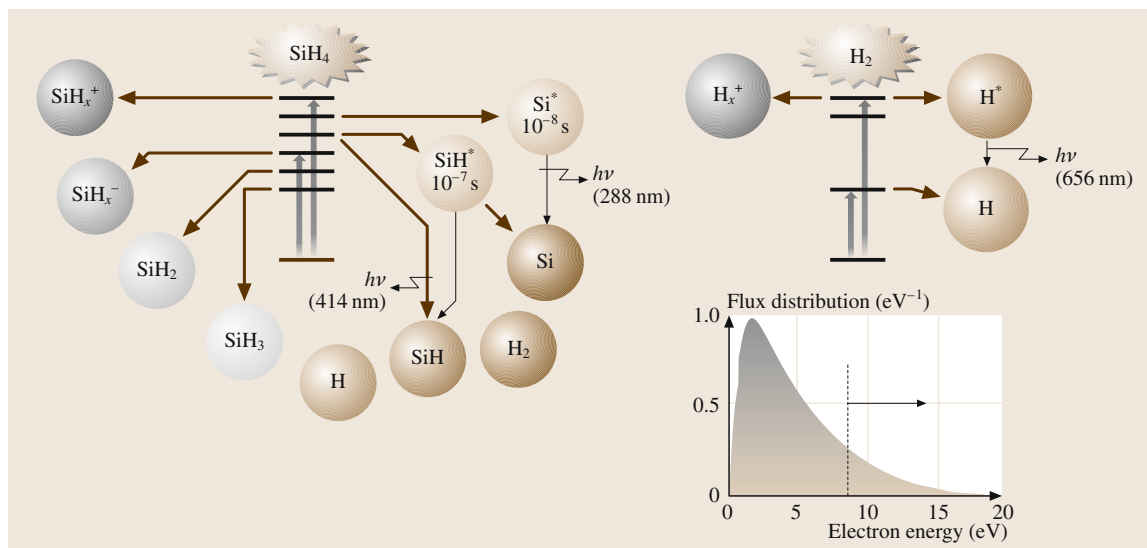


Fig. 26.1 Schematic showing the dissociation of SiH_4 and H_2 molecules to a variety of chemical species in the plasma via excited electronic states. The electron energy distribution function in the plasma is also shown

from zero to several tens of electron volts (eV), ground-state electrons of source gas molecules are excited into their electronic excited states almost simultaneously due to inelastic collisions with energetic electrons. Excited electronic states of complicated molecules like SiH_4 are usually dissociating states, from which dissociation occurs spontaneously to SiH_3 , SiH_2 , SiH , Si , H_2 and H , as shown in Fig. 26.1, depending on the stereochemical structure of the excited state. Hydrogen molecules are also decomposed to atomic hydrogen. Excitation of ground-state electron to vacuum-state gives rise to ionization events, generating new electrons and ions, which maintains the plasma.

Reactive species produced in the plasma also experience secondary reactions, mostly with parent SiH_4 and H_2 molecules, as shown in Fig. 26.2, resulting in a steady state. Reaction rate constants for each reaction are summarized in the literature [26.9]. Steady-state densities of reactive species are basically determined by the balance between their rate of generation and their annihilation rate. Therefore, highly reactive species such as SiH_2 , SiH and Si (short-lifetime species) have much smaller densities than SiH_3 in the steady-state plasma, although the generation rates of those species are not very different from that of SiH_3 , which shows low reactivity along with SiH_4 and H_2 (long-lifetime species).

Steady-state densities of reactive species have been measured using various gas-phase diagnostic techniques [26.10–16], such as optical emission

Ion exchanging	$\text{SiH}_x^+ + \text{SiH}_4 \longrightarrow \text{SiH}_x + \text{SiH}_4^+$
Ion–Molecule	$\text{SiH}_x + \text{SiH}_4 \longrightarrow \text{SiH}_3 + \text{SiH}_3$
Neutral–Molecule	$\text{SiH} + \text{SiH}_4 \longrightarrow \text{Si}_2\text{H}_5$
Disproportionation	$\text{Si} + \text{SiH}_4 \longrightarrow \text{SiH}_3 + \text{SiH}$
Insertion	$\text{SiH}_2 + \text{SiH}_4 \longrightarrow \text{Si}_2\text{H}_6$
Recombination	$\text{SiH}_2 + \text{H}_2 \longrightarrow \text{SiH}_4$
Abstraction	$\text{SiH}_3 + \text{SiH}_4 \longrightarrow \text{SiH}_4 + \text{SiH}_3$
	$\text{H} + \text{SiH}_4 \longrightarrow \text{H}_2 + \text{SiH}_3$
Ion–Radical Radical–Radical	} less probable

Fig. 26.2 Representative secondary reactions of the chemical species produced in the plasma with SiH_4 and H_2 molecules. Their reaction rate constants are available in the literature

spectroscopy (OES) [26.10], laser-induced fluorescence (LIF) [26.12], infrared laser absorption spectroscopy (IRLAS) [26.13], and ultraviolet light absorption spectroscopy (UVLAS) [26.16]. Figure 26.3 shows the steady state number densities of the chemical species, including both emissive and ionic species, in the SiH_4 and SiH_4H_2 plasmas used to prepare device-grade a-Si:H and $\mu\text{c-Si:H}$. It is clear from Fig. 26.3 that the SiH_3 radical is the dominant chemical species in the growth of both a-Si:H and $\mu\text{c-Si:H}$, although the density ratio of short-lifetime species to SiH_3 changes depending on the conditions used to generate the plasma. For instance, when high electric power is supplied to the plasma (high electron density) under low SiH_4 flow

rate conditions, SiH_4 is rapidly dissociated and then depleted, giving rise to reduced probabilities of gas-phase reactions of short lifetime species with SiH_4 molecules. This leads to an increased contribution from short-lifetime species to the film growth, which causes a deterioration of the structural properties of the resulting films.

The steady-state density of atomic hydrogen (H) varies widely in the plasma, as shown in Fig. 26.3. This is mainly due to the change in the hydrogen dilution ratio R (H_2/SiH_4 in the source gas material, i.e., the density of atomic hydrogen increases with increasing R . Noting the fact that more $\mu\text{c-Si:H}$ is formed with increasing R at constant electron density in the plasma and constant substrate (surface) temperature, it is clear

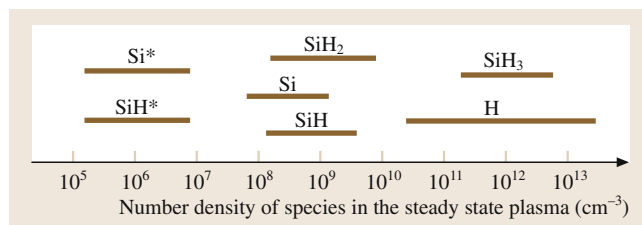


Fig. 26.3 Number densities of chemical species in realistic steady-state plasmas measured or predicted by various diagnostic techniques

that atomic hydrogen plays an important role in $\mu\text{c-Si:H}$ growth [26.17] although SiH_3 is the dominant film precursor for both a-Si:H and $\mu\text{c-Si:H}$ growth [26.18].

26.2 Film Growth on a Surface

26.2.1 Growth of a-Si:H

Upon reaching a film-growing surface, SiH_3 radicals begin to diffuse across it. During this diffusion, the SiH_3 abstracts bonded hydrogen from the surface, forming SiH_4 and leaving dangling bonds on the surface (this known as growth site formation). Other SiH_3 molecules then diffuse across the surface to find the site containing the dangling bonds, whereupon Si–Si bond formation (film growth) occurs, as shown schematically in Fig. 26.4. This surface reaction scheme for film growth has been proposed on the basis of two experimental results [26.18].

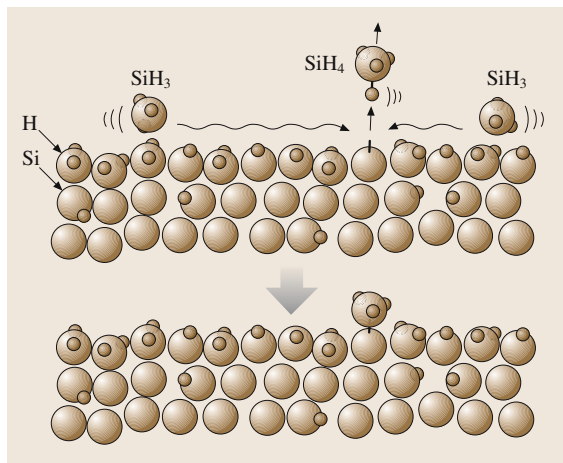


Fig. 26.4 Schematic of the surface growth of a-Si:H

Figure 26.5 shows the general concept of the surface reaction process. Some the flux of SiH_3 is reflected off the surface (the proportion of molecules reflected is given by the reflection probability). The remaining SiH_3 is adsorbed onto the surface and it changes its form as follows:

1. SiH_3 abstracts bonded H from the surface, forming SiH_4 , or two of the SiH_3 radicals interact on the surface, forming Si_2H_6 (with a recombination probability γ);
2. Surface-diffusing SiH_3 sticks to the site containing the dangling bond, forming Si–Si bond (with a sticking probability s).

The total loss probability (β) is given by the sum of recombination probability and the sticking probability ($\gamma + s$), and the reflection probability is therefore $1 - \beta$. Among the reaction probabilities mentioned above, the

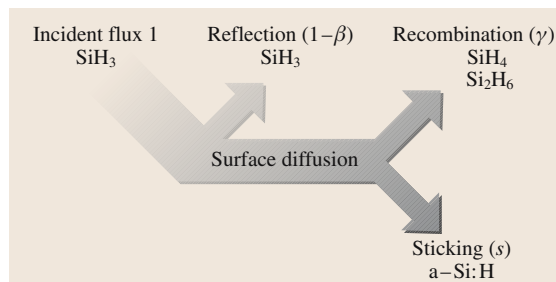


Fig. 26.5 General concepts behind the surface reactions of incoming SiH_3 radicals

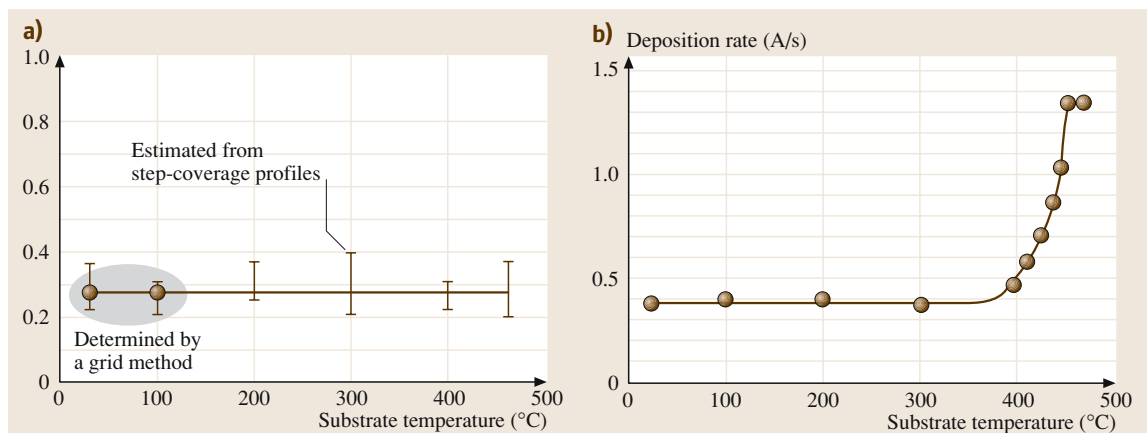


Fig. 26.6a,b Loss probability β (a) of SiH_3 radicals reaching the film-growing surface, as measured using the grid method and the step-coverage method, and the deposition rate (corresponding to the sticking probability s for SiH_3) (b) as a function of substrate temperature during the formation of a-Si:H

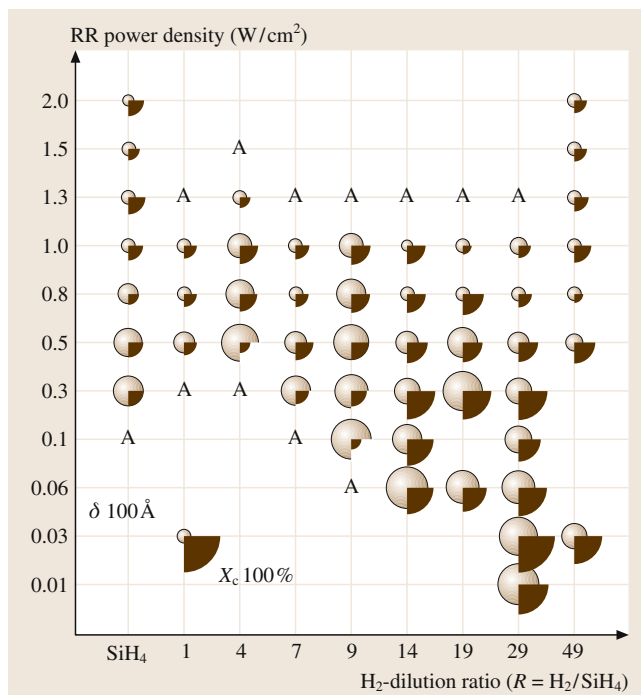


Fig. 26.7 Crystal size (light: δ) and volume % (dark: X_c) of microcrystallites in the resulting films, mapped out on the RF power density/hydrogen dilution ratio plane

loss probability has been measured using a grid method (GM) as well as a step-coverage method (SCM). Figures 26.6a and 26.6b show the loss probabilities obtained by GM and SCM, as well as the sticking probability pre-

dicted from the film deposition rate as a function of the substrate temperature, respectively.

The loss probability β , related to the reflectivity ($1 - \beta$), is simply dependent on the nature of the site where SiH_3 from the plasma lands. Therefore, the temperature-independent $1 - \beta$ seen in Fig. 26.6 suggests that almost all of the surface sites are covered with bonded H over the whole temperature range used in the experiment, although a few dangling bonds are created thermally above 350°C, as shown by in situ infrared reflection absorption spectroscopy (IR-RAS) [26.19]. On the other hand, SiH_3 diffusing across the surface is easily captured by even just a few dangling bonds, because a mobile species on the surface can find a specific site at a distant location from the landing site.

The temperature-independent behavior of $1 - \beta$ and the temperature-dependent behavior of s observed in Figs. 26.6a and 26.6b have been perfectly reproduced by theoretical simulations based on the surface-reaction scheme shown in Fig. 26.4, assuming reasonable activation energies for the surface diffusion of SiH_3 , the abstraction reaction of H with SiH_3 , the saturation of the site containing the dangling bond with SiH_3 , and thermal H-removal processes [26.20].

26.2.2 Growth of $\mu\text{c-Si:H}$

Atomic hydrogen reaching the film-growing surface plays an important role in the growth of $\mu\text{c-Si:H}$. This has been confirmed in the $\mu\text{c-Si:H}$ -formation map draw-

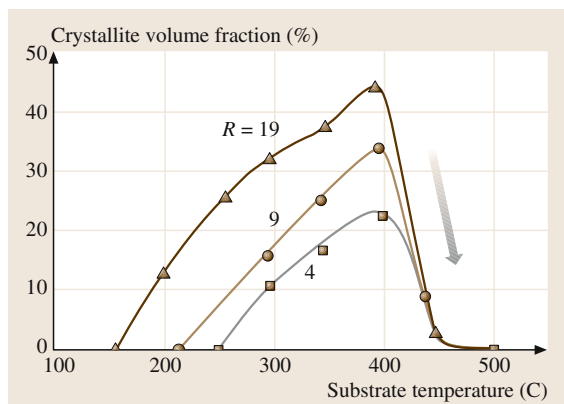


Fig. 26.8 Volume % (X_c) of microcrystallites in the resulting films plotted against the substrate temperature during film growth

ing on the RF power density/hydrogen diffusion ratio plane as shown in Fig. 26.7. As seen in the figure, large-crystallite $\mu\text{c-Si:H}$ is prepared under high hydrogen dilution conditions, indicating the importance of atomic H in the growth of $\mu\text{c-Si:H}$. Figure 26.8 shows the crystalline volume fraction (determined by the X-ray diffraction peak area) in the resulting $\mu\text{c-Si:H}$ as a function of the substrate temperature during film growth for three different hydrogen dilution ratios [26.17]. The crystalline volume fraction increases with increasing substrate temperature, reaching a maximum at around 350 °C, and then suddenly drops to zero above 500 °C. The lack of $\mu\text{c-Si:H}$ formation above 500 °C suggests that surface hydrogen coverage is a requirement for crystallite formation in the resulting film [26.17].

Figure 26.9 shows the negative effect of ionic species impinging on the film-growing surface on the formation of $\mu\text{c-Si:H}$, studied using a triode reactor, as shown in the inset of the figure [26.21]. The Raman crystallinity I_c/I_a , defined as the ratio of the peak intensity from the crystalline phase at around 520 cm^{-1} to that from the amorphous phase at 480 cm^{-1} in the Raman scattering spectrum, deteriorates as the energy of the ionic species impinging on the film-growing surface (controlled by the bias voltage applied to the mesh electrode) is increased.

The surface reaction behavior of SiH_3 reaching the film-growing surface has also been investigated for $\mu\text{c-Si:H}$ growth using GM and SCM, in contrast to a-Si:H growth [26.18]. Figures 26.10a and 26.10b show the loss probability (β) and the deposition rate (corresponding to the sticking probability s) of SiH_3 reaching the

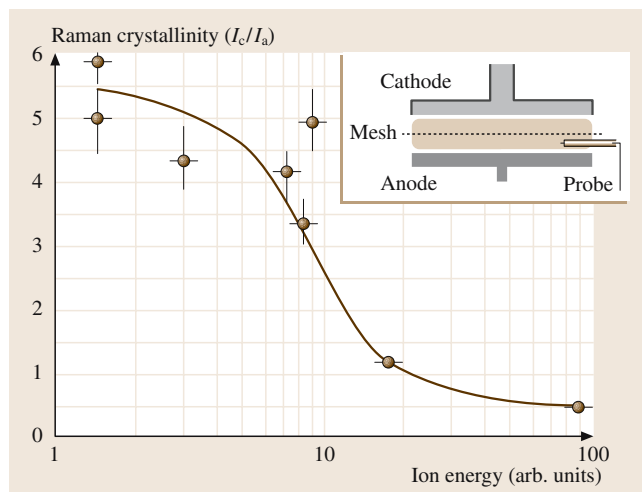


Fig. 26.9 Raman crystallinity in the resulting $\mu\text{c-Si:H}$ as a function of the energy of the ions impinging on the film-growing surface (which is controlled by the application of the bias voltage to the mesh electrode during film growth)

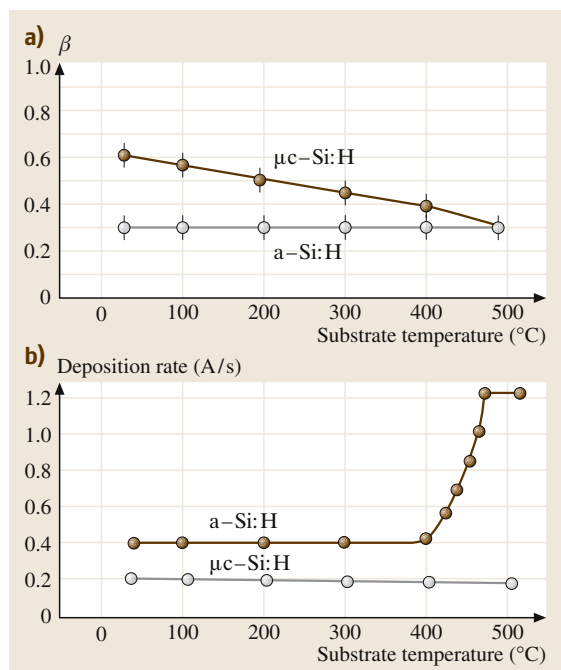


Fig. 26.10 Loss probability β and deposition rate as a function of substrate temperature for $\mu\text{c-Si:H}$ growth in comparison to those for a-Si:H growth

film-growing surface as a function of the substrate temperature for the growth of both $\mu\text{c-Si:H}$ and a-Si:H.

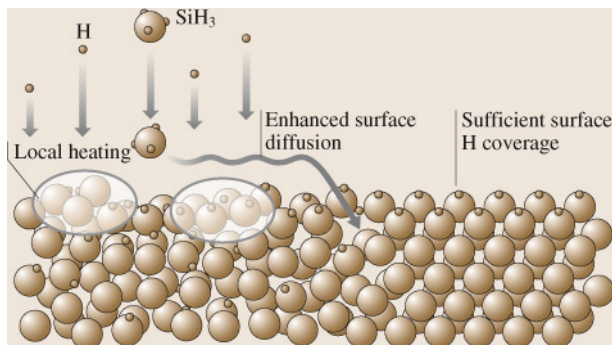


Fig. 26.11 Surface diffusion model for $\mu\text{c-Si:H}$ formation. The large spheres and small spheres represent Si and H, respectively

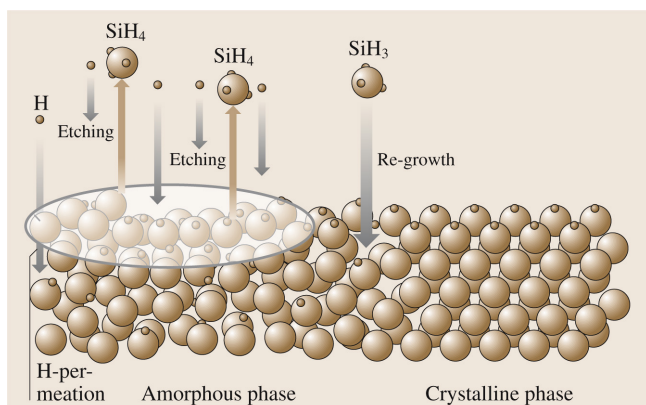


Fig. 26.12 Etching model for $\mu\text{c-Si:H}$ formation

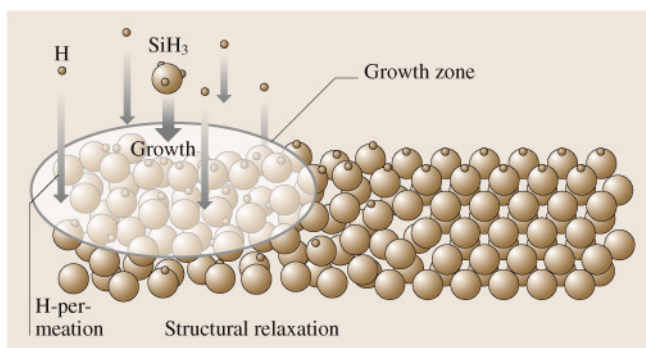


Fig. 26.13 Chemical annealing model for $\mu\text{c-Si:H}$ formation

Unlike in the case of a-Si:H growth, the loss probability β shows a significant dependence on substrate temperature and the sticking probability s shows no temperature dependence in the case of $\mu\text{c-Si:H}$ growth.

Based on these experimental results, the following properties of the formation process of $\mu\text{c-Si:H}$ can be specified:

1. The film precursor is SiH_3 , the same as in the case of a-Si:H growth;
2. Atomic hydrogen reaching the film-growing surface plays an important role in the formation of $\mu\text{c-Si:H}$;
3. The film becomes amorphous when the substrate temperature is higher than 500°C ;
4. High-energy ions impinging on the surface result in crystallinity deterioration;
5. The surface loss probability shows a temperature dependence whereas the sticking probability does not show a temperature dependence in the case of $\mu\text{c-Si:H}$ growth.

Growth Models for $\mu\text{c-Si:H}$

In an attempt to explain the specific phenomena observed during the formation of $\mu\text{c-Si:H}$, three models have been proposed:

1. The surface diffusion model [26.17];
2. The etching model [26.22];
3. The chemical annealing model [26.23].

The surface diffusion model is depicted schematically in Fig. 26.11. Here, a high atomic H flux from the plasma results in full bonded hydrogen surface coverage and also local heating through hydrogen exchange reactions on the film-growing surface. These two actions enhance the surface diffusion of film precursors (SiH_3). As a consequence, the SiH_3 adsorbed on the surface can find energetically favorable sites, leading to the formation of an atomically ordered structure (nucleus formation). After the formation of the nucleus, epitaxial-like crystal growth takes place with enhanced surface diffusion of SiH_3 [26.17, 24].

The etching model has been proposed due to the observation that the rate of film growth decreases with increasing hydrogen dilution ratio R . The concept behind the etching model is shown schematically in Fig. 26.12. Atomic H reaching the film-growing surface breaks Si–Si bonds, preferentially weak bonds involved in the amorphous network structure, leading to the removal of Si atoms weakly bonded to other Si atoms. This site is replaced with a new film precursor SiH_3 , creating a rigid, strong Si–Si bond, which gives rise to an ordered structure [26.22, 24].

The chemical annealing model has been proposed in order to explain the observation that crystal formation is observed during hydrogen plasma treatment;

growth occurs layer-by-layer via an alternating sequence of thin amorphous film growth and hydrogen plasma treatment. Several monolayers of amorphous silicon are deposited, and these layers are exposed to hydrogen atoms produced in the hydrogen plasma. This procedure is repeated several tens of times in order to fabricate the proper thickness to be able to evaluate the film structure. The absence of any significant reduction in film thickness during the hydrogen plasma treatment is difficult to explain using the etching model, and so the chemical annealing model was proposed, as schematically shown in Fig. 26.13. During the hydrogen plasma treatment, many atomic hydrogens permeate through the subsurface (the growth zone), giving rise to the crystallization of an amorphous network through the formation of a flexible network without any significant removal of Si atoms [26.23, 24].

These three models have been carefully examined, and the merits and drawbacks of each model have been discussed [26.24, 25].

More microscopic observations have recently been reported, based on the use of in situ diagnostic techniques, and a detailed mechanism for the formation process of $\mu\text{c-Si:H}$ has been proposed.

Formation of the Nucleus

Figure 26.14 shows the evolution in surface roughness during film growth, as obtained by spectroscopic ellipsometry (SE), for three hydrogen dilution ratios R of 0, 10 and 20 [26.26]. As is seen in the figure, after the formation of an island, the enforced coalescence of islands takes place, which results in a smooth surface under

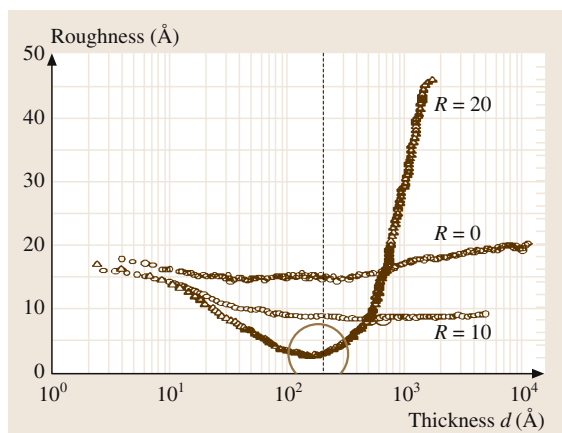


Fig. 26.14 Evolution in surface roughness, as measured using spectroscopic ellipsometry, during film growth for three different R_s

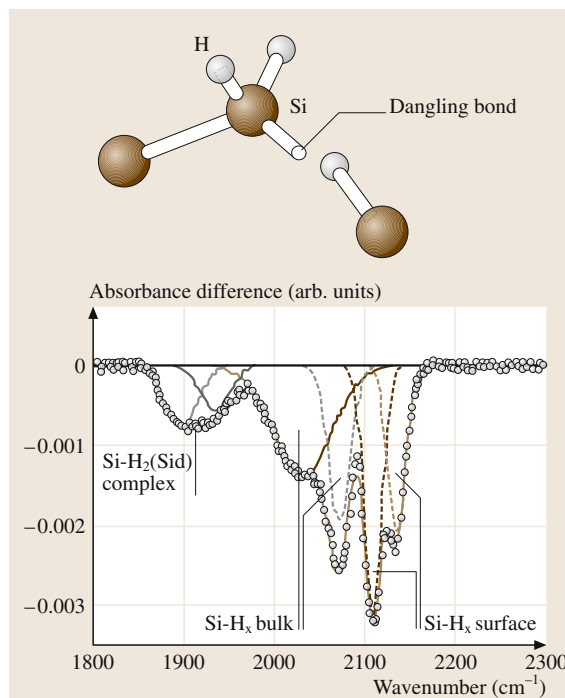


Fig. 26.15 Surface infrared absorption spectrum from the film just before nucleus formation, showing the appearance of the $\text{Si-H}_2\text{-d}$ complex whose structure is also shown

$\mu\text{c-Si:H}$ -growth conditions ($R = 20$). After the smooth surface is obtained (formation of the nucleus is confirmed at this point in time), surface roughness is then enhanced due to an orientation-dependent crystal growth rate [26.27]. As soon as the smooth surface appears, a particular surface absorption band is observed in the infrared absorption spectrum, as measured using the in situ attenuated total reflection technique (ATR) during film growth [26.27, 28]. Figure 26.15 shows the surface infrared absorption spectrum, showing the presence of specific bands at 1897 cm^{-1} and 1937 cm^{-1} together with the usual Si-H_x surface and bulk absorption bands (which occur between 2000 cm^{-1} and 2150 cm^{-1}). This new absorption band is assigned to the $\text{SiH}_2(\text{Sid})$ complex, which is also sketched in Fig. 26.15. Note that the number density (absorption intensity) of the $\text{SiH}_2(\text{Sid})$ complex is found to be proportional to the magnitude of the internal stress in the film just before these complexes appear. A nucleation model has been proposed based on the experimental facts mentioned above. The enforced island coalescence due to the enhanced surface diffusion of SiH_3 gives rise to an internal stress involving many strained Si-Si bonds in the amorphous

incubation layer [26.27]. Atomic hydrogen attacks the strained Si–Si bonds, forming $\text{SiH}_2(\text{Sid})$ complexes on the film-growing surface. These complexes provide structural flexibility, which enables structural order to be obtained via successive Si– SiH_3 bond formation at this site; in other words it acts as a prenucleation site on the film-growing surface.

Epitaxial-Like Crystal Growth

Figure 26.16 shows a cross-sectional transmission electron microscope (TEM) image of typical $\mu\text{c-Si:H}$ films deposited on a glass substrate [26.29]. In the figure, epitaxial-like crystal growth is clearly observed to occur from the nucleus. It is a well-known fact that epitaxial crystal growth occurs only when the surface diffusion length of the film precursor is sufficiently long. In order to investigate the origin of the enhanced surface diffusion of SiH_3 during the formation of $\mu\text{c-Si:H}$, an isotope labeling experiment has been carried out [26.30]. D_2 was used as the source gas material (D_2SiH_4) instead of H_2SiH_4 during film growth under constant substrate temperature conditions, and the number densities of both D and H incorporated into the resulting film are measured by infrared absorption spectroscopy in order to estimate the degree of H-to-D exchange reactions, which mostly occur on the film-growing surface.

Figure 26.17 shows the Raman crystallinity I_c/I_a plotted against the number density ratio of the D/H incorporated in the resulting films. As is clearly from the figure, $\mu\text{c-Si:H}$ is obtained only when D/H exceeds a critical value, indicating that the D/H exchange reaction is required to some extent for the formation of $\mu\text{c-Si:H}$. The D/H exchange reaction involves two steps:

1. Atomic D reaching the film-growing surface abstracts bonded H on the surface (the Eley-Rideal reaction), forming HD and providing a Si dangling bond site;
2. A recombination reaction occurs between the dangling bond and another atomic D adsorbed onto the surface, forming a Si–D bond.

These abstraction and recombination reactions are known to be strong exothermal reactions, with 1.4 eV and 3.1 eV of energy released, respectively. Considering the highly exothermal nature of the D/H exchange reaction, and the existence of a threshold value for the degree of D/H exchange reactions required for the formation

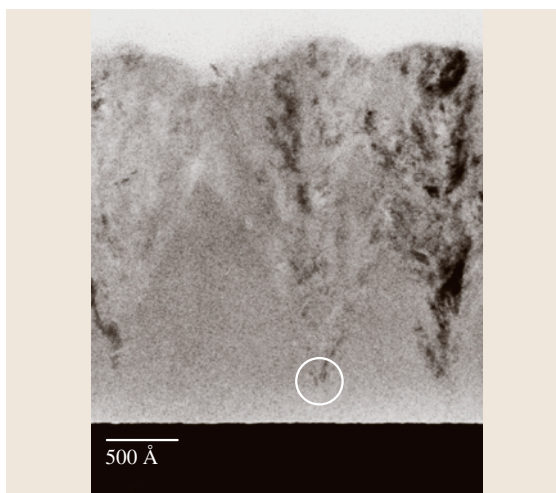


Fig. 26.16 Cross-sectional transmission electron microscope image of typical $\mu\text{c-Si:H}$ deposited on a glass substrate

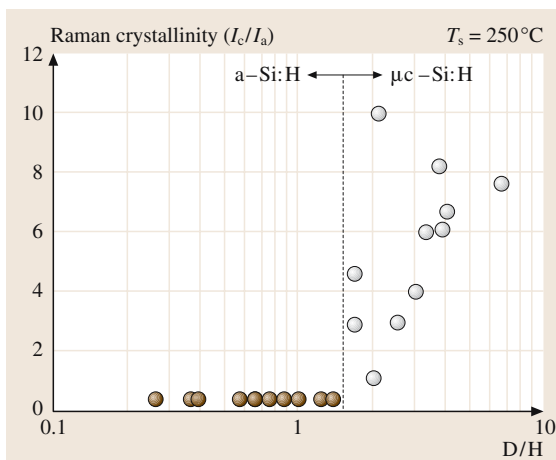


Fig. 26.17 Raman crystallinity plotted against the number density ratio of D/H incorporated into the resulting films prepared from D_2SiH_4 plasmas

of $\mu\text{c-Si:H}$, as shown in Fig. 26.17, “local heating” on the film-growing surface is believed to play a major role in the enhanced surface diffusion of the film precursor (SiH_3) during both nucleus formation (strained bond formation and $\text{SiH}_2(\text{Sid})$ -complex formation) and the epitaxial-like crystal growth associated with $\mu\text{c-Si:H}$ film growth [26.17, 24, 28, 31].

26.3 Defect Density Determination for a-Si:H and $\mu\text{c-Si:H}$

One of the most important structural properties of a-Si:H and $\mu\text{c-Si:H}$ for device applications is their dangling bond defect densities, because each dangling bond creates a localized deep state in the band gap of the material, which acts as a recombination center for photoexcited electrons and holes, although the free carrier mobility is another important property in semiconductors.

26.3.1 Dangling Bond Defects

Figure 26.18 shows the dangling bond defect density in the resulting a-Si:H as a function of substrate temperature. This dependency of the dangling bond density on the substrate temperature has been explained by taking into account the steady-state dangling-bond density on the film-growing surface. The structural properties of the resulting film are generally largely determined by the steady-state surface properties during the thin film growth process, because the surface formed at any given instant is incorporated into the bulk in the next instant due to the successive layering nature of the film growth [26.32, 33]. The steady state number density of surface dangling bonds is determined by the balance between the rate of generation of dangling bonds and their annihilation rate.

At low substrate temperatures, surface dangling bonds are produced by abstraction reactions of surface H atoms with SiH_3 ; the reaction rate for this is almost independent of substrate temperature, since the reaction rate is affected by both the residence time of

SiH_3 at the H-covered site (longer residence times occur at lower temperatures) and by the abstraction reaction rate (slower rates occur at lower temperatures). The surface dangling bond is saturated by SiH_3 diffusing across the surface (this is dangling bond annihilation), which exhibits a slower rate at low temperatures due to the slower surface diffusion of SiH_3 , which results in high dangling bond density on the steady-state film-growing surface. As a consequence, the dangling-bond density in a-Si:H grown at low substrate temperatures can be as high as 10^{19} cm^{-3} . When the substrate temperature is increased during film growth, the steady state number density of surface dangling bonds is reduced drastically due to the combination of the temperature-independent H-abstraction reaction and the thermally enhanced surface diffusion of SiH_3 , which gives rise to a minimum defect density of 10^{15} cm^{-3} in the resulting a-Si:H at a substrate temperature of $\approx 250^\circ\text{C}$. The increased dangling bond density of the a-Si:H prepared at substrate temperatures higher than 350°C is explained by the increased rate of generation of steady-state dangling bonds due to the addition of a new term in the generation rate of these bonds associated with the thermal removal of surface H, although its annihilation rate strictly increases with increasing substrate temperature [26.33, 34].

We note here that a remarkable increase in the contribution from short-lifetime species such as SiH_2 , SiH and Si , which show high reactivity and no diffusion on the film-growing surface, is observed upon the depletion of SiH_4 -parent molecules, which causes an increase in the

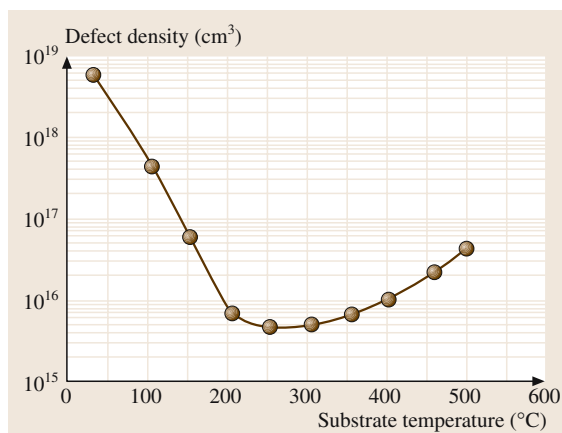


Fig. 26.18 Dangling bond defect density in the resulting a-Si:H films as a function of the substrate temperature during film growth

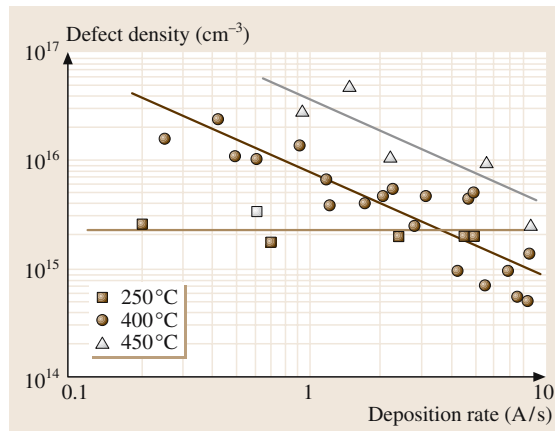


Fig. 26.19 Dangling bond defect density in a-Si:H films plotted against their deposition rates for three different substrate temperatures

number density of dangling bond defects in the resulting films through an enhancement of the dangling bond generation rate and a reduction of the dangling bond annihilation rate on the film-growing surface [26.20,32].

Based on our understanding of the defect density determination reaction during film growth, several trials have attempted to control the defect density in a-Si:H. The steady state defect density on the film-growing surface could be reduced when the growth rate was much faster than the thermal H removal rate in the substrate temperature range above 350 °C, where the steady state defect density is mainly determined by the thermal H removal process. Figure 26.19 shows the number density of dangling bond defects in a-Si:H films plotted against their growth rate. As expected, the defect density in a-Si:H shows no growth-rate dependence when the film is prepared at substrate temperatures lower than 300 °C, whereas the defect density monotonically decreases with increasing growth rate when the film is deposited at 400 °C and 450 °C, and a defect density of 10^{14} cm^{-3} has been demonstrated [26.35].

26.3.2 Dangling Bond Defect Density in $\mu\text{c-Si:H}$

Figure 26.20 shows typical dangling bond defect densities in $\mu\text{c-Si:H}$ as a function of substrate temperature together with those in its a-Si:H counterpart [26.33,35,36]. As seen from the figure, the substrate temperature-dependent dangling bond defect density in $\mu\text{c-Si:H}$ in the temperature range above 300 °C shows a similar trend to that for a-Si:H, indicating that the defect determination reaction on the film-growing surface is identical for both $\mu\text{c-Si:H}$ and a-Si:H growth in this temperature range; in other words, the increase in defect density with increasing substrate temperature is controlled by the thermal

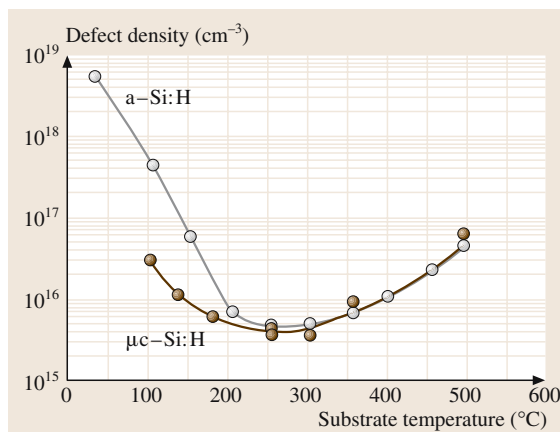


Fig. 26.20 Dangling bond defect density in $\mu\text{c-Si:H}$ films as a function of substrate temperature during film growth in comparison to that in a-Si:H

H removal process on the film-growing surface. On the other hand, the number density of defects in $\mu\text{c-Si:H}$ prepared at low substrate temperatures shows much lower values than those in a-Si:H. Considering that the defect density in the resulting film is largely determined by the defect density on the steady-state film-growing surface, the much lower defect density in $\mu\text{c-Si:H}$ is caused by an increase in the defect annihilation rate on the film-growing surface due to the local heating from H-exchange reactions that occur during the course of $\mu\text{c-Si:H}$ film growth.

It should be noted here that the dangling bond defect density in the resulting $\mu\text{c-Si:H}$ is strongly influenced by the ion bombardment during film growth and by the collisions of crystals growing from different nuclei, which are different aspects to those that are important during a-Si:H growth.

26.4 Device Applications

a-Si:H and $\mu\text{c-Si:H}$ are highly promising materials applicable to electronic or optoelectronic thin-film devices such as thin-film transistors (TFT), position sensors, color sensors, solar cells, etc. [26.1,2].

A thin-film transistor array with a-Si:H active layer has been developed for switching devices in liquid crystal displays (LCD), and large-area LCDs more than 40 in across have already been made commercially available based on this technology. A laser crystallization technique has been developed in order to increase the carrier mobility of a-Si:H-based TFTs and thus reduce the ac-

tive area of the transistors in the LCD. LCDs with a-Si:H or laser-crystallized thin-film Si-based transistor arrays are widely used as flat panel displays in televisions and monitors, where they are in competition with plasma display panels (PDP) and other flat panel display systems. If the carrier mobility in as-deposited $\mu\text{c-Si:H}$ is drastically improved through the enhanced control of the film growth process, $\mu\text{c-Si:H}$ will be widely used not only for thin-film transistors but also for signal-scanning devices such as charge-coupled devices monolithically arranged in the periphery of the LCD.

Thin-film Si-based solar cells has also been widely expected to provide low-cost photovoltaics. Actually, a-Si:H-based solar cells have already been widely used in pocket calculators, and now large-area solar cells are being developed for electricity generation. One big advantage of using a-Si:H for solar cell applications is its large optical absorption coefficient compared to single-crystalline or polycrystalline silicon counterparts, resulting in indirect optical transition properties, and so a thickness of less than $1\ \mu\text{m}$ is enough to absorb sufficient sunlight for electricity generation when using a-Si:H-based solar cells. Low-temperature processes using PECVD are also advantageous in terms of reducing the cost of producing a-Si:H-based solar cells. However, there is a well-known phenomenon that occurs in a-Si:H called photo-induced degradation [26.37], where the conversion efficiency of a-Si:H-based solar cells, usually $\approx 10\%$, is degraded to less than 8% after prolonged exposure to light. Recombinations photo-generated electrons and holes are believed to trigger this photoinduced degradation in a-Si:H; therefore, increasing the field in the a-Si:H-based solar cell has proved an effective way to reduce the degradation. In fact, $0.1\ \mu\text{m}$ -thick a-Si:H-based solar cells do not show

any photoinduced degradation, although the conversion efficiency naturally deteriorates due to the reduced absorption of sunlight. The fabrication of a tandem-type stacked solar cell structure consisting of a top cell with a thin a-Si:H layer and a bottom cell containing narrow-gap materials such as a-SiGe:H, $\mu\text{c-Si:H}$ and $\mu\text{c-SiGe:H}$ has been proposed as a promising way to overcome photoinduced degradation and so to achieve high conversion efficiency in thin-film silicon-based solar cells. Initially, a-SiGe:H was adopted for the bottom cell material [26.38]; however, this material also shows severe photoinduced degradation. Recently, $\mu\text{c-Si:H}$ or $\mu\text{c-SiGe:H}$ have been proposed as promising candidates for bottom cell materials, because these materials do not exhibit photoinduced degradation [26.39,40]. In this proposal, high rates of growth of those materials are crucial to the low-cost fabrication of tandem-type solar cells, since $\mu\text{c-Si:H}$ and $\mu\text{c-SiGe:H}$ undergo largely indirect optical transitions. Therefore, urgent material issues for the realization of low-cost/high-efficiency thin-film silicon-based solar cells include the need to improve the photoinduced stability of high-quality a-Si:H and the need to achieve a high rate of growth of device-grade $\mu\text{c-Si:H}$ or $\mu\text{c-SiGe:H}$.

26.5 Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells

26.5.1 Controlling Photoinduced Degradation in a-Si:H

A relationship between the degree of photoinduced degradation and the dihydride bonding (Si-H_2) density has been reported in a-Si:H prepared under a variety of deposition conditions, where the substrate temperature, plasma-excitation frequency, gas-flow rate, hydrogen dilution ratio, working pressure, power density, etc. have all been varied [26.41]. Figure 26.21 shows the degree of photoinduced degradation, defined as the difference in the fill factors of photo-I-V characteristics of Ni-a-Si:H Schottky diode before and after light soaking plotted

against the Si-H_2 density in a-Si:H film, as measured by infrared absorption spectroscopy. Furthermore, it has been suggested from mass spectrometric results that the Si-H_2 density in the resulting a-Si:H is strongly in-

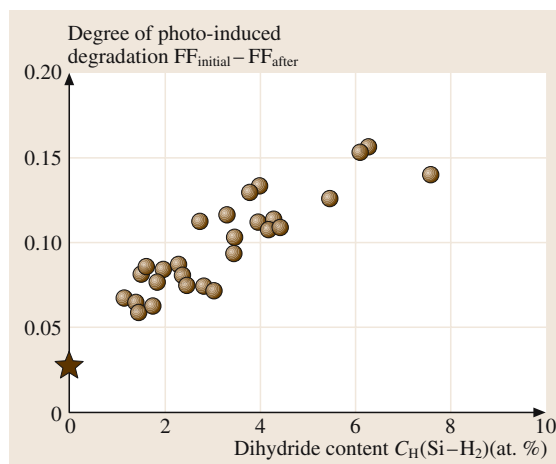


Fig. 26.21 Relationship between the degree of photoinduced degradation and the dihydride (Si-H_2) content in a-Si:H prepared under various deposition conditions. The star symbol represents for a-Si:H prepared under conditions of a reduced ratio of contributions from higher silane-related species versus the SiH_3 contribution at a substrate temperature of 250°C

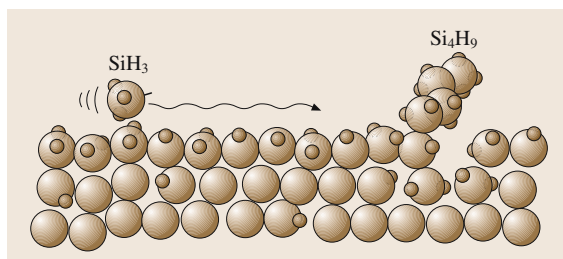


Fig. 26.22 Surface reaction image of the incorporation of Si–H₂ bonds into the resulting a-Si:H due to the contributions from higher silane-related species such as Si₄H₉

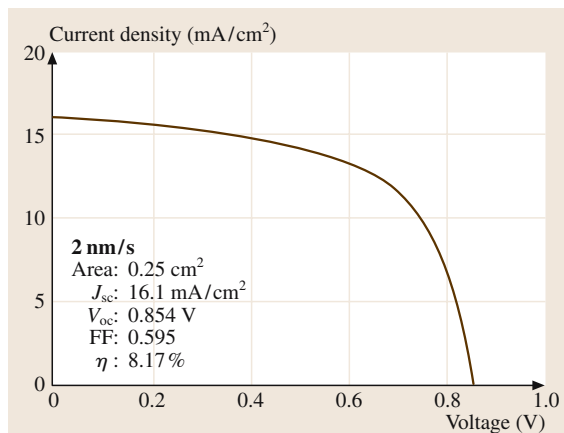


Fig. 26.23 Photo-I–V characteristics of a stable n–i–p a-Si:H-based solar cell fabricated at a high growth rate of 2 nm/s

creased by the contributions from higher silane-related species (HSRS) such as Si₄H₉ when the substrate temperature is kept constant. The contribution of these HSRS to the film precursor (SiH₃) has been theoretically analyzed using a couple of gas-phase reaction-rate equations [26.42]. This analysis predicted that the contribution ratio is a complex function of the electron temperature in the plasma, the electron density in the plasma, the gas temperature, the hydrogen dilution ratio R , and the gas residence time during film growth. Based on our understanding of the network structure (Si–H₂ bonding configuration) responsible for the photoinduced degradation in a-Si:H and the chemical species (HSRS) responsible during film growth (shown schematically in Fig. 26.22), a guiding principle for obtaining highly stabilized a-Si:H has been proposed [26.42, 43].

By following the guiding principle, a-Si:H with minimized Si–H₂ density in the network has been prepared by adjusting the plasma parameters during film growth

under high growth rate conditions, and an a-Si:H-based solar cell showing a stable conversion efficiency of 8.2% (Fig. 26.23) has been fabricated at a high growth rate of 2 nm/s [26.41].

Moreover, a-Si:H containing a Si–H₂ density of 0% has been successfully prepared using a triode reactor at substrate temperatures as low as 250 °C by making use of the difference in the gas phase diffusion coefficients of SiH₃ (light) and HSRS (heavy) during film growth.

26.5.2 High Growth Rates of Device-Grade μ c-Si:H

In the past, conventional high hydrogen dilution methods performed at relatively low working gas pressures (several tens of mTorr) have been used to obtain device-grade μ c-Si:H (with low dangling bond defect densities of $\approx 10^{16}/\text{cm}^3$) [26.17]. Recently, a simple concept for preparing device-grade μ c-Si:H at a high growth rate has been proposed, known as the high-pressure depletion (HPD) method [26.21, 44]. The production rate of the film precursor SiH₃ in the plasma, which is proportional to the growth rate of μ c-Si:H, is determined by the product of the electron density and the number density of SiH₄. To increase the production rate of SiH₃ in the plasma, a high power density (radio frequency, RF, or very high frequency, VHF) and high partial pressures are needed, because the electron density is basically a function of the power density applied

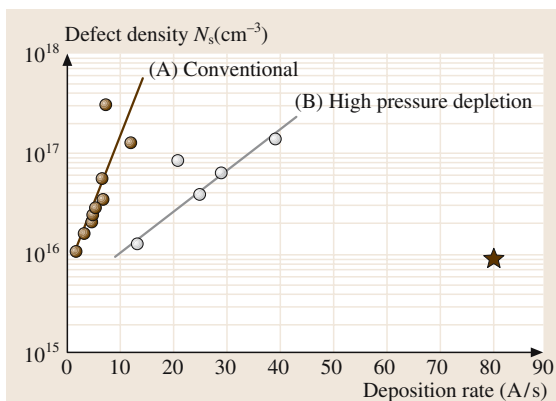


Fig. 26.24 Relationships between the dangling bond defect density in μ c-Si:H films and their deposition rates for (A) μ c-Si:H prepared using the conventional low-pressure regime and for (B) μ c-Si:H prepared using the high-pressure depletion (HPD) method. The star symbol represents μ c-Si:H prepared under HPD conditions with the novel cathode design

to the plasma and the number density of SiH_4 is proportional to the partial pressure. However, hydrogen atoms (the chemical species responsible during the formation of $\mu\text{-Si:H}$) are strongly scavenged by SiH_4 molecules during their transportation from their production site to the film-growing surface. In order to enhance the survival of the hydrogen atoms, it has been suggested that the SiH_4 molecules should be depleted by applying high power density to the plasma. As conditions of high total pressure are also useful for decreasing the effects of ion bombardment during film growth through the reduction of the electron temperature in the plasma, high working pressures along with SiH_4 depletion conditions (HPD) have recently been popularly adopted for the high-rate growth of device-grade $\mu\text{-Si:H}$.

Figure 26.24 shows the number density of dangling bond defects, as measured by electron spin resonance (ESR), in the resulting $\mu\text{-Si:H}$ as a function of growth rate [26.45]. When the conventional low pressure regime is used for the growth of $\mu\text{-Si:H}$, the defect density increases exponentially with increasing growth rate, as seen in the figure (see A). This is caused by both an increase in the ion bombardment (due to the high power density) and an increase in the contributions from short-lifetime chemical species such as SiH_2 , SiH and Si due to the reduced ability of those species to react with SiH_4 and H_2 (due to the low pressure and SiH_4 depletion) during film growth. On the other hand, the slope in Fig. 26.24 becomes shallower when HPD conditions are used during the growth of $\mu\text{-Si:H}$, as shown in the figure (see B), which illustrates the usefulness of the HPD method for obtaining high-quality $\mu\text{-Si:H}$ at high growth rates.

The validity of the HPD method has also been demonstrated during the fabrication of $\mu\text{-Si:H}$ -based solar cells. Figure 26.25 shows the photo-I-V characteristics of a $\mu\text{-Si:H}$ -based p-i-n single-junction solar cell prepared at a high growth rate of 2 nm/s using the HPD method, which exhibits a reasonably high conversion efficiency of 8.1% [26.46].

However, HPD conditions require that the spacing between the cathode and anode is reduced in the conventional capacitively coupled plasma reactor, giving rise to nonuniform plasma production and nonuniform film growth. To overcome this problem encountered when using HPD conditions, the structure of the cathode surface has been designed to produce uniform plasma

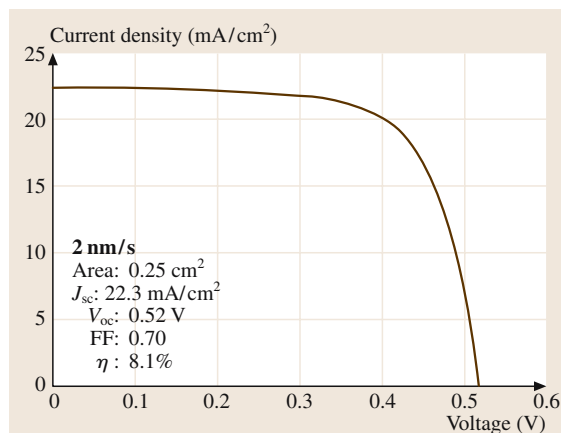


Fig. 26.25 Photo-I-V characteristics of a p-i-n $\mu\text{-Si:H}$ -based solar cell fabricated at reasonably high growth rate of 2 nm/s under HPD conditions

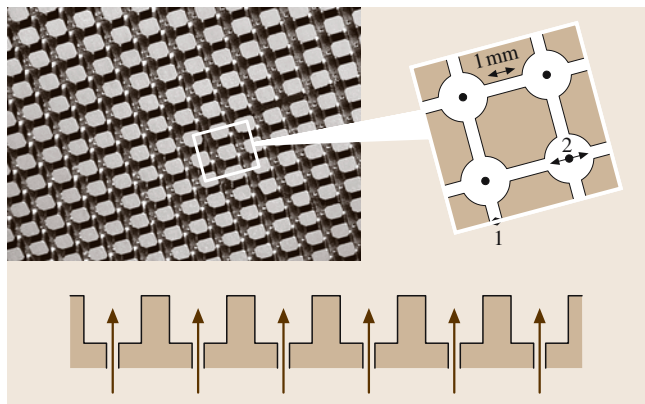


Fig. 26.26 Photograph of the surface structure on the cathode designed for the production of high-density/uniform plasmas

production even in a large-area parallel plate electrode configuration. Figure 26.26 shows the structure of this novel design of cathode. A multitude of holes (hollows) with interconnecting slots are arranged on the cathode surface, which cause strong coupling between the high-density plasmas produced in each hole where source gas injection is performed. Using this new type of cathode, quite high growth rates (more than 8 nm/s) have been obtained along with reasonably low defect densities in the resulting $\mu\text{-Si:H}$, as shown by star symbol in Fig. 26.24 [26.47].

26.6 Summary

In this chapter, the processes involved in the growth of a-Si:H and μ c-Si:H from SiH₄ and SiH₄H₂ plasma have been interpreted in detail. The defect density determination reaction that takes place on the film-growing surface was discussed in order to obtain clues that may lead to enhanced optoelectronic properties in those materials. The recent status of work done in the fields of thin-film transistors and solar cells was reviewed, as these are

the main device applications of those materials. Recent progress in resolving material issues related to solar cell applications were also described.

Finally, we note here that the concepts used in and our understanding of the film growth process mentioned here are widely applicable to other processes, especially processes where thin films are grown from reactive plasmas.

References

- 26.1 W. E. Spear, P. G. LeComber: Solid State Commun. **17**, 1193 (1975)
- 26.2 C. R. Wronski, D. E. Carlson, R. E. Daniel: Appl. Phys. Lett. **29**, 602 (1976)
- 26.3 T. Moustakas: *Semicond. Semimet.*, Vol. 21A (Academic, New York 1984) p. 55
- 26.4 T. Saito, S. Muramatsu, T. Shimada, M. Migitaka: Appl. Phys. Lett. **42**, 678 (1983)
- 26.5 Y. Mishima, M. Hirose, Y. Osaka, K. Nagamine, Y. Ashida, K. Isogaya: Jpn. J. Appl. Phys. **22**, L46 (1983)
- 26.6 T. Fuyuki, K. Y. Du, S. Okamoto, S. Yasuda, T. Kimoto, M. Yoshimoto, H. Matsunami: J. Appl. Phys. **64**, 2380 (1988)
- 26.7 A. H. Mahan, B. P. Nelson, S. Salamon, R. S. Crandall: Mater. Res. Soc. Proc. **219**, 673 (1991)
- 26.8 M. Tsuda, S. Oikawa, K. Saito: J. Chem. Phys. **91**, 6822 (1989)
- 26.9 J. Perrin, O. Leroy, M. C. Bordage: Contrib. Plasma Phys. **36**, 3 (1996)
- 26.10 A. Matsuda, K. Nakagawa, K. Tanaka, M. Matsumura, S. Yamasaki, H. Okushi, S. Iizima: J. Non-Cryst. Solids **35–36**, 183 (1980)
- 26.11 A. Matsuda, K. Tanaka: Thin Solid Films **92**, 171 (1982)
- 26.12 Y. Matsumi, T. Hayashi, H. Yoshikawa, S. Komiya: J. Vac. Sci. Technol. A **4**, 1786 (1986)
- 26.13 N. Itabashi, N. Nishiwaki, M. Magane, T. Goto, A. Matsuda, C. Yamada, E. Hirota: Jpn. J. Appl. Phys. **29**, 585 (1990)
- 26.14 N. Itabashi, N. Nishiwaki, M. Magane, S. Saito, T. Goto, A. Matsuda, C. Yamada, E. Hirota: Jpn. J. Appl. Phys. **29**, L505 (1990)
- 26.15 K. Tachibana, T. Mukai, H. Harima: Jpn. J. Appl. Phys. **30**, L1208 (1991)
- 26.16 A. Kono, N. Koike, H. Nomura, T. Goto: Jpn. J. Appl. Phys. **34**, 307 (1995)
- 26.17 A. Matsuda: J. Non-Cryst. Solids **59–60**, 767 (1983)
- 26.18 A. Matsuda, T. Goto: Mater. Res. Soc. Proc. **164**, 3 (1990)
- 26.19 Y. Toyoshima, K. Arai, A. Matsuda, K. Tanaka: J. Non-Cryst. Solids **137–138**, 765 (1991)
- 26.20 J. L. Guizot, K. Nomoto, A. Matsuda: Surf. Sci. **244**, 22 (1991)
- 26.21 M. Kondo, M. Fukawa, L. Guo, A. Matsuda: J. Non-Cryst. Solids **266–269**, 84 (2000)
- 26.22 C. C. Tsai, G. B. Anderson, R. Thompson, B. Wacker: J. Non-Cryst. Solids **114**, 151 (1989)
- 26.23 K. Nakamura, K. Yoshida, S. Takeoka, I. Shimizu: Jpn. J. Appl. Phys. **34**, 442 (1995)
- 26.24 A. Matsuda: Thin Solid Films **337**, 1 (1999)
- 26.25 K. Saito, M. Kondo, M. Fukawa, T. Nishimiya, W. Futako, I. Shimizu, A. Matsuda: Res. Soc. Proc. Mater. **507**, 843 (1998)
- 26.26 J. Koh, Y. Lee, H. Fujiwara, C. R. Wronski, R. W. Collins: Appl. Phys. Lett. **73**, 1526 (1998)
- 26.27 H. Fujiwara, M. Kondo, A. Matsuda: Surf. Sci. **497**, 333 (2002)
- 26.28 H. Fujiwara, Y. Toyoshima, M. Kondo, A. Matsuda: J. Non-Cryst. Solids **266–269**, 38 (2000)
- 26.29 H. Fujiwara, M. Kondo, A. Matsuda: Phys. Rev. B **63**, 115306 (2001)
- 26.30 S. Suzuki, M. Kondo, A. Matsuda: J. Non-Cryst. Solids **299–302**, 93 (2002)
- 26.31 H. Fujiwara, M. Kondo, A. Matsuda: Jpn. J. Appl. Phys. **41**, 2821 (2002)
- 26.32 A. Matsuda, K. Nomoto, Y. Takeuchi, A. Suzuki, A. Yuuki, J. Perrin: Surf. Sci. **227**, 50 (1990)
- 26.33 G. Ganguly, A. Matsuda: Phys. Rev. B **47**, 3361 (1993)
- 26.34 G. Ganguly, A. Matsuda: J. Non-Cryst. Solids **164–166**, 31 (1993)
- 26.35 G. Ganguly, A. Matsuda: Jpn. J. Appl. Phys. **31**, L1269 (1992)
- 26.36 Y. Nasuno, M. Kondo, A. Matsuda: Tech. Digest of PVSEC-12. Jeju, Korea, (2001) 791
- 26.37 D. L. Staebler, C. R. Wronski: Appl. Phys. Lett. **28**, 671 (1977)
- 26.38 Ke. Saito, M. Sano, K. Matsuda, T. Kondo, T. Nishimoto, K. Ogawa, I. Kajita: Proc. WCPEC-2 Vienna, Austria (1998) p.351
- 26.39 J. Meier, P. Torres, R. Platz, S. Dubail, U. Kroll, J. A. Anna Selvan, N. Pellaton Vaucher, Ch. Hof, D. Fischer, H. Keppner, A. Shah, K. -D. Ufort, P. Giannoulas: Mater. Res. Soc. Proc. **420**, 3 (1996)

- 26.40 K. Yamamoto, M. Yoshimi, T. Suzuki, Yu. Tawada, Y. Okamoto, A. Nakajima: *Mater. Res. Soc. Proc.* **507**, 131 (1998)
- 26.41 T. Nishimoto, M. Takai, H. Miyahara, M. Kondo, A. Matsuda: *J. Non-Cryst. Solids* **299-302**, 1116 (2002)
- 26.42 M. Takai, T. Nishimoto, T. Takagi, M. Kondo, A. Matsuda: *J. Non-Cryst. Solids* **266-269**, 90 (2000)
- 26.43 M. Takai, T. Nishimoto, M. Kondo, A. Matsuda: *Appl. Phys. Lett.* **77**, 2828 (2000)
- 26.44 L. Guo, M. Kondo, M. Fukawa, K. Saito, A. Matsuda: *Jpn. J. Appl. Phys.* **37**, L1116 (1998)
- 26.45 M. Kondo, T. Nishimoto, M. Takai, S. Suzuki, Y. Nasuno, A. Matsuda: *Tech. Digest of PVSEC-12 Jeju, Korea* (2001) 41
- 26.46 T. Matsui, M. Kondo, A. Matsuda: *Proc. WCPEC-3 Osaka, Japan* (2003) 50-A3-02
- 26.47 C. Niikura, M. Kondo, A. Matsuda: *Proc. WCPEC-3 Osaka, Japan* (2003) 5P-D4-03

Ferroelectric

27. Ferroelectric Materials

Ferroelectric materials offer a wide range of useful properties. These include ferroelectric hysteresis (used in nonvolatile memories), high permittivities (used in capacitors), high piezoelectric effects (used in sensors, actuators and resonant wave devices such as radio-frequency filters), high pyroelectric coefficients (used in infra-red detectors), strong electro-optic effects (used in optical switches) and anomalous temperature coefficients of resistivity (used in electric-motor overload-protection circuits). In addition, ferroelectrics can be made in a wide variety of forms, including ceramics, single crystals, polymers and thin films – increasing their exploitability. This chapter gives an account of the basic theories behind the ferroelectric effect and the main ferroelectric material classes, discussing how their properties are related to their composition and the different ways they are made. Finally, it reviews the major applications for this class of materials, relating the ways in which their key functional properties

27.0.1	Definitions and Background	597
27.0.2	Basic Ferroelectric Characteristics and Models	599
27.1	Ferroelectric Materials	601
27.1.1	Ferroelectric Oxides	601
27.1.2	Triglycine Sulphate (TGS)	607
27.1.3	Polymeric Ferroelectrics	607
27.2	Ferroelectric Materials Fabrication Technology	608
27.2.1	Single Crystals	608
27.2.2	Ceramics	609
27.2.3	Thick Films	613
27.2.4	Thin Films	613
27.3	Ferroelectric Applications	616
27.3.1	Dielectrics	616
27.3.2	Computer Memories	616
27.3.3	Piezoelectrics	617
27.3.4	Pyroelectrics	620
	References	622

affect those of the devices in which they are exploited.

27.0.1 Definitions and Background

Ferroelectric materials offer a very wide range of useful properties for the electronic engineer to exploit. As we will see, they are also a class of materials that is hard to define accurately in a single sentence. It is useful to start from the class of insulating materials that form dielectrics; in other words materials that will sustain a dielectric polarisation under the application of an electric field. There exists a set of these materials for which the crystal structure lacks a centre of symmetry. (If a crystal structure has a centre of symmetry, it means that for every atom in the structure there is a point in the unit cell through which inversion will bring one to the same type of atom.) A list of the non-centrosymmetric, or *acentric*, point groups is given in Table 27.1. All of the crystalline materials whose structures possess these point groups (with the exception of group 432) exhibit the phenomenon of piezoelectricity, which means that stress will generate a charge separation on the faces of

the crystal (the direct piezoelectric effect) and will undergo mechanical strain when subjected to an electric field (the converse piezoelectric effect). Both effects are widely exploited in electronic devices. A well-known example of a non-centrosymmetric material is the mineral α -quartz, which is used for the piezoelectric resonators employed for frequency filtering and electronic clocks.

Table 27.1 Polar and acentric (non-centrosymmetric) point groups

Crystal system	Polar (acentric)	Nonpolar (acentric)
Triclinic	1	
Monoclinic	2, m	
Orthorhombic	$mm2$	222
Trigonal	3, $3m$	32
Hexagonal	6, $6mm$	$\bar{6}$, $\bar{6}m2$
Tetragonal	4, $4mm$	$\bar{4}$, 422, $\bar{4}2m$
Cubic	None	23, $\bar{4}3m$, 432

A sub-set of the non-centrosymmetric crystals also possess a unique axis of symmetry. These crystals are said to be *polar*. The polar point groups are also listed in Table 27.1. Polar crystals are piezoelectric (as they are acentric) and also exhibit pyroelectricity, which means that a charge separation will appear on their surfaces when their temperature is changed. Polar structures effectively have a dielectric polarisation *built in* to the unit cell of the crystal structure. This is sometimes called a *spontaneous polarisation*. The application of stress or a change in temperature causes a change in this dipole moment and it is this change that causes the separation of charge on the surfaces of the crystal. The direction and magnitude of the spontaneous polarisation in a polar dielectric can be changed by the application of an electric field, but on removal of the field it will return to its zero-field value. A well-known example of a polar dielectric is ZnO, which possesses the wurtzite crystal structure in which Zn^{2+} ions sit in the tetrahedrally coordinated sites between hexagonal close-packed layers of oxygen ions. Thin films of ZnO are widely used for piezoelectric applications.

A sub-set of the set of polar dielectrics exists for which the application of a field of sufficient magnitude will cause the spontaneous polarisation to switch to a different, stable direction. Upon removal of the field the polarisation will not spontaneously return to its original direction and magnitude. These crystals are called *ferroelectric*. We can view the relationships between the sets of ferroelectric, polar, acentric and centrosymmetric dielectrics as shown in the Venn diagram in Fig. 27.1.

The history of ferroelectrics is long, as has been described in the excellent review by Busch [27.1]. Many very distinguished scientists were involved in its early development, including Brewster (who was one of those who studied pyroelectricity), J. & P. Curie (who discovered piezoelectricity), Boltzmann, Pockels and Debye,

to name but a few. Indeed, the term ferroelectricity was first coined by E. Schrodinger. However, credit for the discovery of the effect goes to Joesph Valasek, who found in 1920 that the polarisation of sodium potassium tartrate (Rochelle salt) could be switched by the application of an electric field, thus providing the first demonstration of the process that is the hallmark of ferroelectricity.

A reasonable working definition of ferroelectricity is “a polar dielectric in which the polarisation can be switched between two or more stable states by the application of an electric field”. However, as we will see in the discussion which follows, there are exceptions to this definition: some ferroelectrics are semiconducting (and thus are not dielectrics because they cannot sustain an electrical polarisation); the spontaneous polarisation in some ferroelectrics cannot be switched because they cannot sustain an electric field of sufficient magnitude to effect the switching, either because they reach electrical breakdown first, or because they are too conducting.

Since the initial discovery of ferroelectricity in Rochelle salt, the effect has been demonstrated in a wide range of materials, from water-soluble crystals through oxides to polymers, ceramics and even liquid crystals. Many of these will be discussed in this chapter. The range of useful properties exhibited by ferroelectrics covers:

- Ferroelectric hysteresis is used in nonvolatile computer information storage.
- Ferroelectrics can exhibit very high relative permittivities (several thousand) which means that they are widely used in capacitors.
- The direct piezoelectric effect (the generation of charge in response to an applied stress) is widely used in sensors such as accelerometers, microphones, hydrophones etc.
- The converse piezoelectric effect (the generation of strain in response to an applied electric field) is widely used in actuators, ultrasonic generators, resonators, filters etc.
- The pyroelectric effect (the generation of charge in response to a change in material temperature) is widely used in uncooled infra-red detectors.
- The electro-optic effect (a change in birefringence in response to an applied electric field) is used in laser Q-switches, optical shutters and integrated optical (photonic) devices.
- Ferroelectrics exhibit strong nonlinear optical effects that can be used for laser frequency doubling and optical mixing.

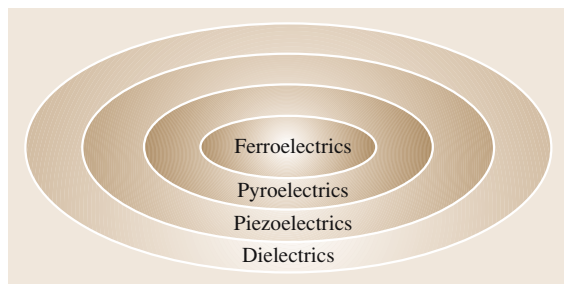


Fig. 27.1 Venn diagram showing how ferroelectrics fit into the different classes of dielectric materials

- Illumination of transparent ferroelectrics with light of sufficient energy causes excitation of carriers into the conduction band. Their movement under the internal bias field caused by the spontaneous polarisation causes a refractive-index modulation that can be used for a variety of optical applications, including four-wave mixing and holographic information storage.
- Ferroelectrics exhibit strong coupling between stress and birefringence, which can be used to couple acoustic waves to optical signals with applications in, for example, radar signal processing.
- Doping certain ferroelectric ceramics with electron donors (e.g. BaTiO₃ with La³⁺) can render them semiconducting. Heating these ceramics through their Curie temperature causes a very large, reversible increase in resistivity (by several orders of magnitude in some cases) over a narrow range of temperature (ca. 10 °C). This large positive temperature coefficient of resistance (PTCR) is widely exploited in electric-motor overload-protection devices and self-stabilising ceramic heating elements.

27.0.2 Basic Ferroelectric Characteristics and Models

The switchable spontaneous polarisation in ferroelectrics gives rise to the first characteristic property of the materials: ferroelectric hysteresis. Figure 27.2 shows a schematic plot of polarisation versus electric field as would be observed on a typical ferroelectric. As the field is increased from zero, the overall polarisation in the crystal increases as the polarisations in different dipolar regions are aligned. Eventually, it reaches a saturation

point where the only further increase in P is that due to the relative permittivity of the material. (The gradient of the P/E curve for a linear dielectric is equal to its permittivity). Extrapolation of this line back to the abscissa gives the saturation value of the spontaneous polarisation (P_s). Reduction of the field to zero leaves a *remanent polarisation* (P_r), which is usually slightly less than P_s . A negative field will cause the polarisation to reduce, until it reaches zero at the *coercive field* ($-E_c$). A further negative increase in the field will eventually cause a reverse saturation polarisation ($-P_s$) to develop. When the field returns to zero the crystal is left with a negative remanent polarisation ($-P_r$). Increasing the field once more, increases the polarisation from $-P_r$ to zero at E_c , and then to $+P_s$, completing the ferroelectric hysteresis loop. The ability to switch the polarisation between two states gives rise to the first application of ferroelectrics – as a nonvolatile memory storage medium. It also permits polycrystalline ferroelectrics (especially ceramics) to be polarised. The fact that the polarisation can possess different directions within different regions of the same crystal gives rise to the existence of *ferroelectric domains*. These are given names according to the angle between the polarisation between adjacent regions. Hence, two adjacent regions in which the polarisations are orientated at 180° to each other are, naturally, called 180° domains. In many ferroelectrics, the polarisation can choose between many different possible directions so that adjacent regions can have polarisations at, for example, 90° to one another.

For most ferroelectrics, the polar state only exists over a limited range of temperatures. As the temperature is raised, a point is reached at which there is a transition from the polar, ferroelectric phase to a non-polar, non-ferroelectric phase (called the *paraelectric* phase). In all cases, the paraelectric phase possesses a higher crystal symmetry than the ferroelectric phase into which it transforms. The temperature at which this occurs is called the *Curie temperature* (T_C). This ferroelectric-to-paraelectric phase transition is a characteristic of most ferroelectrics, but again there are exceptions. Some ferroelectric materials melt or decompose before T_C is reached; the polymeric ferroelectric polyvinylidene fluoride (PVDF) is one such example. For most ferroelectrics (called *proper ferroelectrics*), as the Curie temperature is approached from above, the relative permittivity is observed to increase, reaching a peak at T_C and decreasing below T_C . (There is a class of ferroelectrics, known as *improper* or *extrinsic ferroelectrics*, for which there is no peak in permittivity, just an anomaly. These are described further below.) Very

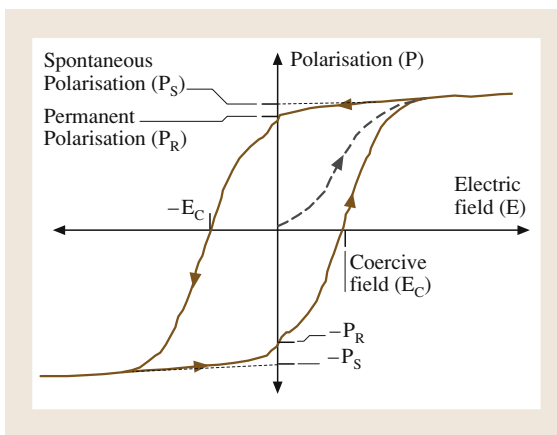


Fig. 27.2 Ferroelectric hysteresis loop

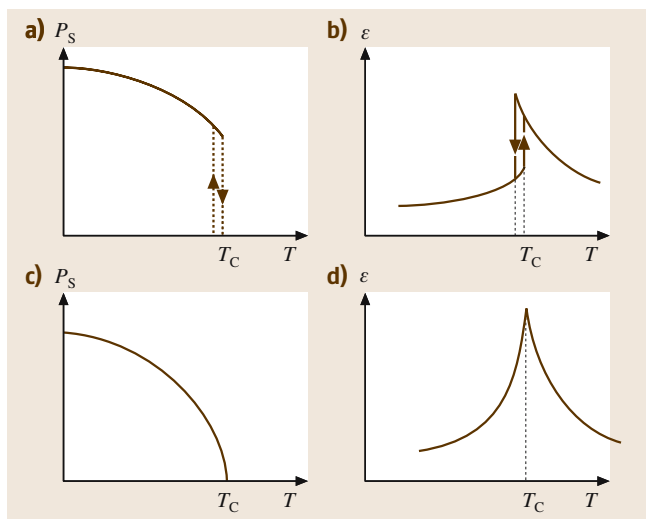


Fig. 27.3a–d The variations of spontaneous polarisation (a and c) and permittivity (b and d) with temperature for typical ferroelectric materials. (a) and (b) are for a first-order (discontinuous) phase transition while (c) and (d) are for a second-order (continuous) phase transition

high values of relative permittivity (many thousands) can be reached at T_C , leading to another use of ferroelectric materials – as dielectrics in high-value capacitors. The phase transition at T_C can either be continuous (known as second-order) or discontinuous (known as first-order). In either case, the dependence of permittivity (ϵ) on temperature above T_C can be described by $\epsilon = C/(T - T_0)$. T_0 is called the *Curie Weiss temperature*, and is only equal to T_C for a second-order phase transition. C is called the *Curie Weiss constant*. Figure 27.3 shows typical plots of P_s and ϵ as functions of temperature for ferroelectrics around T_C . Figs. 27.3a and 27.3b show the behaviour for a first-order phase transition. As can be seen, P_s drops discontinuously to zero at the phase-transition temperature. The permittivity rises as the temperature decreases, peaking at the transition, where there is a discontinuous step down. First-order phase transitions tend to show thermal hysteresis in the transitions, so that the transition occurs at a higher temperature when approached from the low-temperature side than from the high-temperature side. Frequently in practical observations there is actually a region of coexistence where both phases exist together in the same sample at the same temperature. In this case, the Curie temperature is defined as the temperature at which the ferroelectric and paraelectric phases have the same Gibbs free energies. In the case of a second-order phase transition (Figs. 27.3c

and 27.3d), P_s drops continuously to zero at T_C and the permittivity rises to a sharp peak. There is no hysteresis in the transition.

There are two types of theoretical model that describe the ferroelectric phenomenon. The first is the phenomenological or thermodynamic theory first proposed by *Devonshire* [27.2]. Here, the Gibbs free energy of the ferroelectric system is described in terms of a power series in the spontaneous polarisation. In this case, P_s is termed the order parameter for the phase transition and the theory is similar in form to the generalised phenomenological theory of phase transitions known as Landau theory. It is not appropriate to go into the details of the theory here, which has been described very well by *Lines and Glass* [27.3]. The theory successfully predicts many of the behavioural characteristics of ferroelectric materials in terms of the measurable macroscopic properties, including the ferroelectric hysteresis and the behaviour of the permittivity near T_C . However, it tells us nothing about the microscopic origins of ferroelectricity. This is the province of the two types of microscopic models of ferroelectric behaviour. The first of these is the order–disorder model of ferroelectric behaviour. The second is the displacive model.

According to the order–disorder model, the electric dipoles exist within the structure in the paraelectric phase above T_C , but are thermally disordered between two or more states so that the average polarisation is zero. There are several materials for which that is the case, an example being potassium dihydrogen phosphate (KDP), which is paraelectric, tetragonal ($\bar{4}2m$) above 123 K and ferroelectric, orthorhombic ($mm2$) below. The polarisation appears along the tetragonal c -axis. In this crystal structure, the PO_4 groups form tetrahedra that are linked by hydrogen bonds at their corners. The protons in these sit in double potential wells. Above T_C they are delocalised between the two minima in the wells. Below T_C , they are localised and the PO_4 tetrahedra become distorted. The phosphorous and potassium ions become displaced relative to the oxygen framework, forming the polarisation in the lattice. Replacement of the hydrogen in the structure by deuterium raises T_C to about 220 K because the heavier deuteron delocalises between the two potential minima at a much higher temperature. Another example of an order–disorder ferroelectric is sodium nitrite ($NaNO_2$) in which the polar groups are NO_2^- ions, which become ordered below T_C (163 °C) with their dipoles all pointing along the b -axis of the crystal structure.

In displacive ferroelectrics, there are considered to be no dipoles in the structure above T_C , but the dipoles

appear in the structure due to the cooperative motion of ions. Typically this is seen as the *softening* of a zone-centre optical phonon. What this means is that the frequency of an optical phonon mode with zero wavevector goes to zero at T_C . Such a phonon mode involves the displacement in opposite directions of cations and anions in the structure. When the frequency of this mode goes to zero, a dipolar displacement results. Examples of such ferroelectrics include the perovskite-structured oxides such as BaTiO_3 , which will be discussed further below. One of the successes of soft-mode theory is that it provides a convincing mechanism for the peak

in the dielectric permittivity at T_C . According to the Lyddane–Sachs–Teller relationship (see, for example, [27.3] p. 216), the softening of a zone-centre phonon will make the permittivity diverge at T_C .

In practice, the distinction between order–disorder and displacive ferroelectrics becomes rather blurred, as some ferroelectrics which may be considered to be purely displacive may exhibit significant disordered cation displacement well above T_C , while some order–disorder ferroelectrics show soft-mode behaviour. For a more detailed discussion of these theories, the reader is referred to *Lines and Glass* [27.3].

27.1 Ferroelectric Materials

27.1.1 Ferroelectric Oxides

This is by far the most technologically important class of ferroelectric materials. They will be discussed according to the most important crystal classes.

Perovskite Ferroelectrics

These materials possess crystal structures isomorphous with the mineral perovskite (CaTiO_3). They all have the general chemical formula ABO_3 , where A and B are cations. Typically the A cation will be around 1.2–1.6 Å in radius (similar to the oxygen ions) while the B cations

will be around 0.6–0.7 Å in radius. The crystal structure is illustrated in Fig. 27.4. It consists of a network of corner-linked BO_6 octahedra, within which is enclosed the large A cation. Another way to look at the structure is as cubic-close-packed AO_3 layers, with the small B cations sitting in the octahedral sites between these close-packed layers. A pair of layers are shown in Fig. 27.5. The structure is a very tolerant one and will accommodate many different ions. Because of this, it is exhibited by a large number of oxides. The basic criteria for the structure to be stable is that the valencies of the ions should balance and that the ionic radii meet the *Goldschmidt criteria* [27.4]. A tolerance factor t is

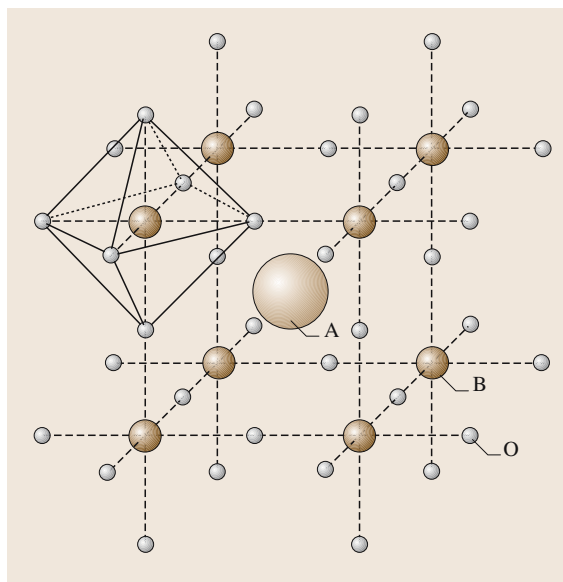


Fig. 27.4 The perovskite crystal structure

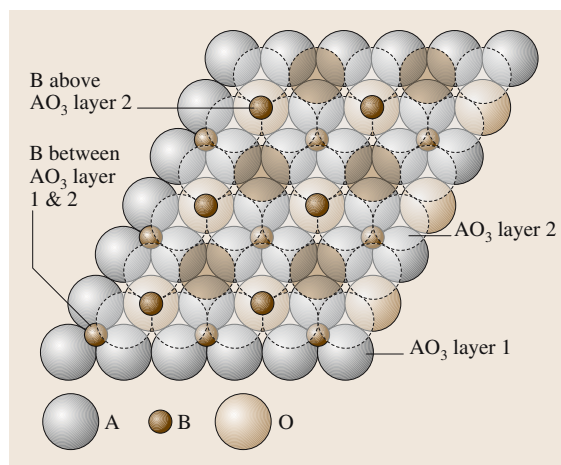


Fig. 27.5 Showing two AO_3 close-packed layers of the perovskite structure, with the B cations sitting in the sixfold coordinated sites between the layers. The second layer has been made semitransparent for clarity

Table 27.2 List of cations frequently found to form perovskite-structured oxides and their ionic radii. Ionic radii taken from [27.5]

A-site cation	Ionic radius when [12] ^a by O ²⁻ (Å)	B-site cation	Ionic radius when [6] ^b by O ²⁻ (Å)
Na ⁺	1.32	Nb ⁵⁺	0.64
K ⁺	1.6	Ta ⁵⁺	0.68
Ba ²⁺	1.6	Zr ⁴⁺	0.72
Sr ²⁺	1.44	Ti ⁴⁺	0.605
Pb ²⁺	1.49	Pb ⁴⁺	0.775
Bi ³⁺	1.11	Sc ³⁺	0.73
Ca ²⁺	1.35	Fe ³⁺	0.645

^a “12-fold coordinated”; ^b “6-fold coordinated”

defined:

$$t = \frac{r_A + r_O}{\sqrt{2}(r_B + r_O)},$$

where r_X is the ionic radius of the X cation. The ideal cubic perovskite structure, where the ions are just touching each other, will possess $t = 1$. However, the structure will be stable with $0.85 < t < 1.05$. A list of ions that commonly form perovskites, together with their ionic radii, is given in Table 27.2. The closer t is to unity, the more likely the structure will be to be cubic. Conversely, perovskites which have values of $t < 1$ show distorted structures that are frequently ferroelectric. Examples of some perovskites and their tolerance factors are listed in Table 27.3. This table also lists the structures formed by the compounds at room temperature, and whether or not they are ferroelectric. Some of the most interesting perovskites from the point of view of applications are BaTiO₃, PbTiO₃ and KNbO₃. BaTiO₃ is cubic above 135 °C, but transforms to a tetragonal fer-

roelectric structure below this temperature. In this case, the Ba and Ti ions are displaced relative to the anion framework along one of the cubic <001> directions. This means that the polar axis has six choices for direction in the tetragonal phase. At 5 °C there is a second-phase transition from the tetragonal to an orthorhombic phase, where the polarisation now appears due to cation displacements along one of the cubic <110> directions, for which there are 12 choices. Finally, at –90 °C there is a transition to a rhombohedral phase with the cations being displaced along one of the cubic <111> directions, for which there are eight choices. In the case of PbTiO₃ there is a single transition to a tetragonal phase at 490 °C, again by cationic displacements along <100>. Some perovskites show phase transitions that are not ferroelectric. For example, SrTiO₃ shows a transition to a tetragonal phase at 110 K which involves linked rotations, or tilts, of the TiO₆ octahedra about the cubic [100] direction. Tilting of the octahedra is a common feature of the phase transitions that occur in perovskites and can lead to very complex series of phase transitions, as has been observed for NaNbO₃. This type of structural modification and a commonly-used notation for it has been described in detail by Glazer [1972].

It is very easy to make solid solutions of the end-member perovskites, such as those listed in Table 27.3, and this has been used to great effect to provide materials with a wide range of properties. For example, Fig. 27.6 shows the temperature dependence of the relative permittivity of a BaTiO₃ single crystal [27.6]. There is a peak at each transition where the value perpendicular to the polar axis reaches several thousand, making the material interesting for use in capacitor dielectrics. However, the temperature variation in such a material would make it useless. The formation of ceramic solid solutions of BaTiO₃ with SrTiO₃, CaTiO₃ or PbTiO₃ allows the temperature dependence of permittivity to be controlled

Table 27.3 Some end-member perovskites and their properties

Perovskite oxide	Tolerance factor	Structure at 20 °C	Type	T _C (°C)
BaTiO ₃	1.06	Tetragonal	Ferroelectric	135
SrTiO ₃	1.00	Cubic	Paraelectric	
CaTiO ₃	0.97	Tetragonal	Paraelectric	
PbTiO ₃	1.02	Tetragonal	Ferroelectric	490
PbZrO ₃	0.96	Orthorhombic	Antiferroelectric	235
NaNbO ₃	0.94	Monoclinic	Ferroelectric	–200
KNbO ₃	1.04	Tetragonal	Ferroelectric	412
KTaO ₃	1.02	Cubic	Ferroelectric	–260
BiScO ₃	0.83	Rhombohedral	Ferroelectric	370
BiFeO ₃	0.87	Tetragonal	Ferroelectric	850

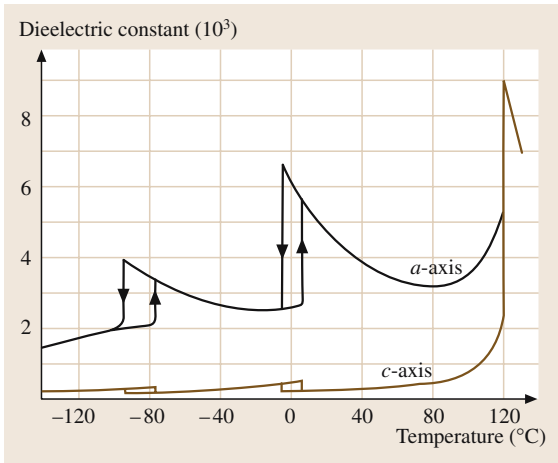
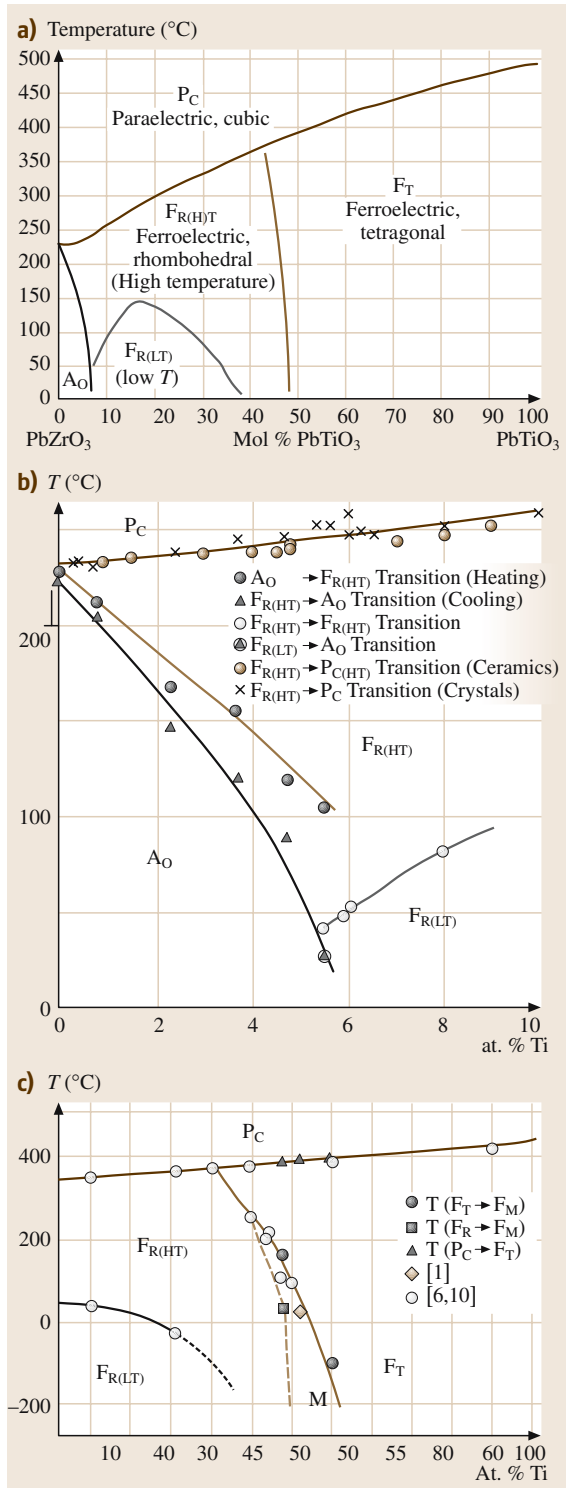


Fig. 27.6 Temperature dependence of the relative permittivity of BaTiO₃ measured along [001] and [100] (After [27.6])

so that useful specifications can be met with average permittivities of 2000 or more over a wide range of temperatures. *Herbert* [27.7] has discussed the details of these modifications and their effects on the BaTiO₃ transition temperatures. BaTiO₃-based ceramics are widely used in ceramic capacitors and form the basis of an industry worth billions of dollars annually. They have also been used for piezoelectric ceramics, but are no longer so important in that field, as they have largely been displaced by ceramics in the PbZrO₃–PbTiO₃ system. Nevertheless, there is a renewed interest in BaTiO₃-based ceramics for lead-free piezoelectrics in response to the legislative drive to reduce lead in the environment.

The perovskite solid solution system between PbZrO₃ and PbTiO₃ is of great technological importance and is thus worth discussing in some detail. The phase diagram given in Fig. 27.7 shows several phases. Starting at PbTiO₃, the ferroelectric tetragonal (F_T) phase persists well across the diagram, until the composition Pb(Zr_{0.53}Ti_{0.47})O₃ is reached, where there is a transition to a ferroelectric, rhombohedral phase. The composition where this occurs is called the morphotropic phase boundary (MPB). The rhombohedral phase region splits into two. There is a high-temperature phase (F_{R(HT)}) in which the cations are displaced along the cubic [111] direction. There is also a low-temperature ferroelectric rhombohedral (F_{R(LT)}) phase in which the (Zr, Ti)O₆

Fig. 27.7 (a) PbZrO₃–PbTiO₃ phase diagram (after [27.8]) (b) Region close to PbZrO₃ (after [27.9, 10]) (c) Region close to the MPB (after [27.11])



octahedra are rotated about the [111] axis. This doubles the unit cell. Close to PbZrO_3 (Fig. 27.7b) the room-temperature structure transforms to the antiferroelectric, orthorhombic (A_0) structure. This is a complex structure in which the cations are displaced in double antiparallel rows along the cubic (110) directions, coupled with octahedral tilts. A higher-temperature A_T phase, which is frequently shown when this phase diagram is cited in the literature, has been shown not to exist in pure solid solutions [27.10]. There is another, more modern, modification to this phase diagram. It has been shown by *Noheda et al.* [27.11] that a monoclinic phase exists at the MPB (Fig. 27.7c). The enormous technological and commercial importance of this system derives from the very high piezoelectric, pyroelectric and electro-optic coefficients that can be obtained from ceramic compositions in different parts of the phase diagram, particularly when the base composition is doped with selected ions. These are discussed in more detail below. This solid solution system is frequently referred to generically as *PZT* although, strictly, *PZT* was the brand name for a set of piezoelectric ceramic compositions manufactured by the Clevite Corporation.

There is another very important class of oxides, termed complex perovskites, where the A or, more commonly B, sites in the structure are occupied by ions of different valency in a fixed molar ratio. These are not solid solutions as they possess a fixed composition. Examples are $\text{PbMg}_{1/3}\text{Nb}_{2/3}\text{O}_3$ (PMN), $\text{PbSc}_{1/2}\text{Ta}_{1/2}\text{O}_3$ (PST) and $\text{Bi}_{1/2}\text{Na}_{1/2}\text{TiO}_3$. These are also technologically important and can form end members to solid solutions in their own right, $\text{PbMg}_{1/3}\text{Nb}_{2/3}\text{O}_3$ – PbTiO_3 (PMN-PT), for example. Many of the complex ferroelectrics (PMN-PT is a classic example) exhibit *ferroelectric relaxor* behaviour. In these materials, the Curie point is no longer a sharp transition but is actually observed over a very wide range of temperatures. A broad permittivity peak is observed and the temperature and height of the peak are strongly dependent on the frequency of measurement. Figure 27.8 shows a typical variation of permittivity with temperature and frequency in PMN [27.12].

The broad tolerance of the perovskite structure to different cations has led to a wide exploration of the inclusion of different iso-valent dopants to obtain different electronic properties. It is also possible to include alio-valent dopants over a wide range, which the structure tolerates through the introduction of cation or anion vacancies, or free charge carriers. For example, it is possible to incorporate Nb^{5+} into the B site of the BaTiO_3 system [27.13], and charge balance is main-

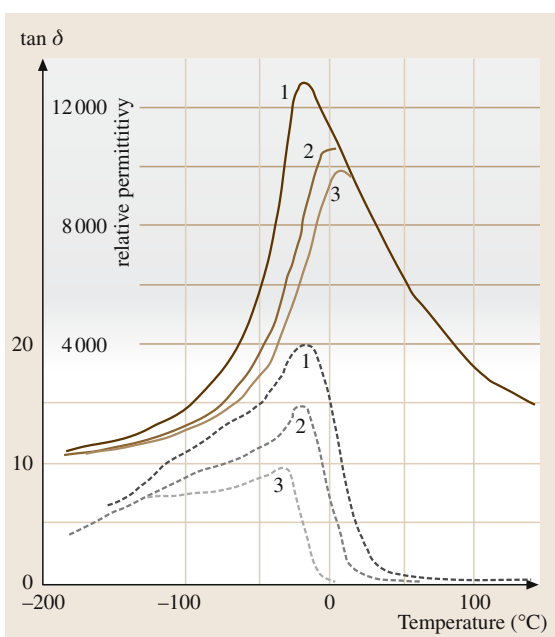


Fig. 27.8 Permittivity (solid lines) and dielectric loss (dashed lines) vs. temperature in PMN. The measurements were made at: (1) 0.4, (2) 45 and (3) 4500 kHz. (Adapted from [27.12])

tained through the presence of free electrons. Similarly, Fe^{3+} can be introduced into the B site of the system and this substitution creates O vacancies [27.14].

The vast majority of perovskites are employed in ceramic form, but thin- and thick-film materials are becoming of increasing importance and single-crystal materials are starting to emerge which have exceptionally high piezoelectric coefficients.

Illmenite Ferroelectrics

The illmenite structure is related to the perovskite structure in that it is exhibited by materials with the general formula ABO_3 , where the A cation is too small to fill the [12] coordinated site of the perovskite structure. The structure is made up of hexagonal close-packed layers of oxygen ions, with the A and B ions occupying the octahedrally coordinated sites between the layers. Hence, this structure can also be considered to be related to the perovskite structure in that both are based on oxygen octahedra. The two best known illmenite ferroelectrics are LiNbO_3 and LiTaO_3 , whose structure is shown in Fig. 27.9. The materials have high T_C values (ca. 1200 °C and 620 °C, respectively). In the case of LiNbO_3 , the T_C is only 50 °C below the melting point.

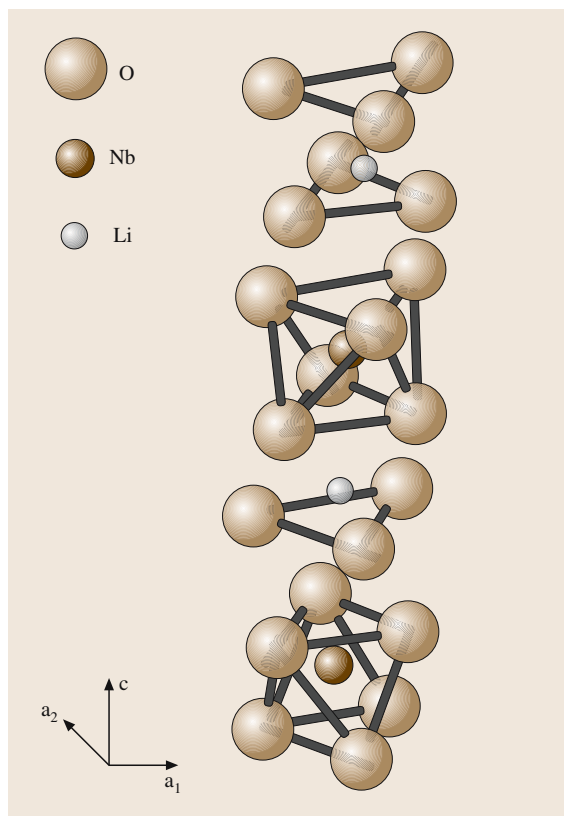


Fig. 27.9 Crystal structure of LiNbO_3 (after [27.15])

Referring to Fig. 27.9, it can be seen that the cations occupy octahedral sites progressing along the c -axis in the order Nb, vacancy, Li, Nb, vacancy, Li etc. The polarisation occurs through displacement of the cations along the three-fold axis, so the structure transforms from $\bar{3}m$ to $3m$ symmetry at T_C . These are *uniaxial* ferroelectrics, in that the polarisation can only be up or down, and only 180° domains exist. The materials are mainly used in single-crystal form for piezoelectric and electro-optic devices.

Tungsten Bronze Ferroelectrics

This is another very large family of oxygen octahedral ferroelectrics possessing the general formula $[\text{A}_1\text{A}_2\text{A}_4\text{C}_4][\text{B}_1\text{B}_2\text{B}_8]\text{O}_{30}$. The crystal structure is complex, shown schematically in Fig. 27.10. The B1 and B2 sites are octahedrally coordinated by oxygens and have similar sizes and valencies to the B sites in the perovskites. The A1 and A2 sites are surrounded by four and five columns of BO_6 octahedra respectively. The three-fold coordinated C sites in the structure are

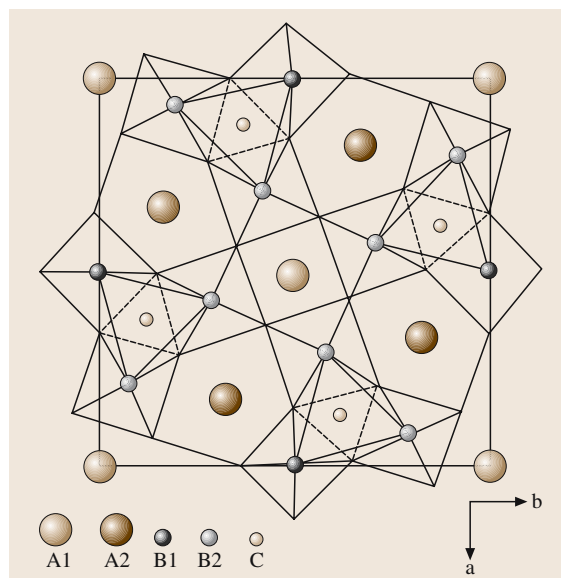


Fig. 27.10 Tungsten bronze crystal structure (after [27.16])

frequently empty, but can be occupied by small uni- or divalent cations (e.g. Li^+ or Mg^{2+}). There are a wide range of ferroelectric tungsten bronzes, which frequently show nonstoichiometry. All are tetragonal in their paraelectric phase, and can transform to either a tetragonal ferroelectric phase, in which the polar axis appears along the tetrad axis of the paraelectric phase, or (more commonly) into an orthorhombic phase in which the polar axis appears perpendicular to the original tetrad axis. Examples of tungsten bronzes are PbNb_2O_6 (lead metaniobate) in which five out of the available six A sites are occupied by Pb^{2+} and the B sites by Nb^{5+} . This crystal and its Ta analogue are metastable below about 1200°C . They can be stabilised by rapid cooling, or more commonly by doping. Lead metaniobate ceramics are difficult to make, but are commercially available and are occasionally used for piezoelectric devices by virtue of their higher Curie temperatures than members of the PZT system, combined with reduced lateral piezoelectric coupling factors. Substitution of Pb^{2+} by Ba^{2+} can also be used to produce a useful piezoelectric ceramic material $\text{Pb}_{1/2}\text{Ba}_{1/2}\text{Nb}_2\text{O}_6$. Other tungsten bronze ferroelectrics such as $\text{Sr}_x\text{Ba}_{1-x}\text{Nb}_2\text{O}_6$, $\text{Ba}_2\text{NaNb}_5\text{O}_{15}$ and $\text{K}_3\text{LiNb}_5\text{O}_{15}$ have been investigated as single crystals for use in electro-optic devices. However, the growth of highly homogeneous single crystals is difficult because of the formation of optical striations caused by compositional fluctuations.

Aurivillius Compounds

The aurivillius compounds form another important class of ferroelectrics based on oxygen octahedra. They have a good deal in common with the perovskites in that they consist of layers or slabs of perovskite blocks with the general formula $(A_{m-1}B_mO_{3m+1})^{2-}$ separated by $(M_2O_2)^{2+}$ layers in which the M cation is in pyramidal coordination with four oxygens, the M being at the apex of the pyramid. The structure possesses the general formula $M_2(A_{m-1}B_mO_{3m+3})$. In the ferroelectric phases, M is usually Bi and m is usually between one and five. The A and B cations follow the usual ionic radius and valency criteria of the perovskites, and we see ferroelectric compounds in which A = Bi, La, Sr, Ba, Na, K etc. and B = Fe, Ti, Nb, Ta, etc. Part of the crystal structure is shown schematically in Fig. 27.11, illustrating in this case a structure with $m = 3$. This is just over half of the unit cell, showing one slab of perovskite units and two of the $(Bi_2O_2)^{2-}$ layers. The structures are tetragonal in their paraelectric phases, with the tetrad c -

axis being perpendicular to the $(Bi_2O_2)^{2+}$ plane. In the majority of cases they become orthorhombic in their ferroelectric phase, with the spontaneous polarisation appearing in the $a-b$ plane, at 45° to the tetragonal a -axis. Bismuth titanate ($Bi_4Ti_3O_{12}$) is an example of an aurivillius ferroelectric in which $M = A = Bi$, $B = Ti$ and $m = 3$, so that there are three blocks in the perovskite slab. In this case, there is a small component of the spontaneous polarisation (ca. $4 \mu C/cm^2$) out of the $a-b$ plane, with the majority of the polarisation (ca. $50 \mu C/cm^2$) being in the plane, so that the ferroelectric structure is monoclinic. The T_C is high ($675^\circ C$) and ceramics have been explored for use as high-temperature piezoelectrics. The problem with the compounds when used as ceramics is that the piezoelectric coefficients are rather small. Single crystals have been explored for optical light-valve applications but have not received widespread use because of the problems of growth. Thin films of $SrBi_2Ta_2O_9$ (SBT) are receiving considerable attention for applications in ferroelectric nonvolatile memories.

Other Oxide Ferroelectrics

There is a wide range of other oxide ferroelectric materials, although none have anywhere near the breadth of application of the oxygen octahedral ferroelectrics discussed above. There are two groups of phosphates (in which the phosphorous ions are tetrahedrally coordinated by oxygens) which have found applications in optical systems. The potassium dihydrogen phosphate (KDP) family was referred to above as an example of an order-disorder ferroelectric. KDP, and its deuterated analogue KD^*P , can be grown from aqueous solution as large, high-quality single crystals. These can be lapped and polished into plates which are used for longitudinal electro-optic modulators, in which the modulating electric field is applied parallel to the direction of light propagation in the crystal. In this application, the crystals are used at room temperature in the paraelectric phase. One spectacular application for these crystals is as the Q-switches in the large ultra-high-power lasers used for nuclear fusion research, where extremely large (ca. 1 m) crystal slices are needed. Another class of phosphate ferroelectrics is the potassium titanyl phosphate system (KTP). These crystals can be grown relatively easily from the melt and are used for optical frequency doubling applications, particularly using periodically poled crystals [27.17, 18]. This is a large family in which the potassium can be replaced by caesium or rubidium and the phosphorous by arsenic. Lasers using KTP are finding significant uses in surgical procedures [27.19].

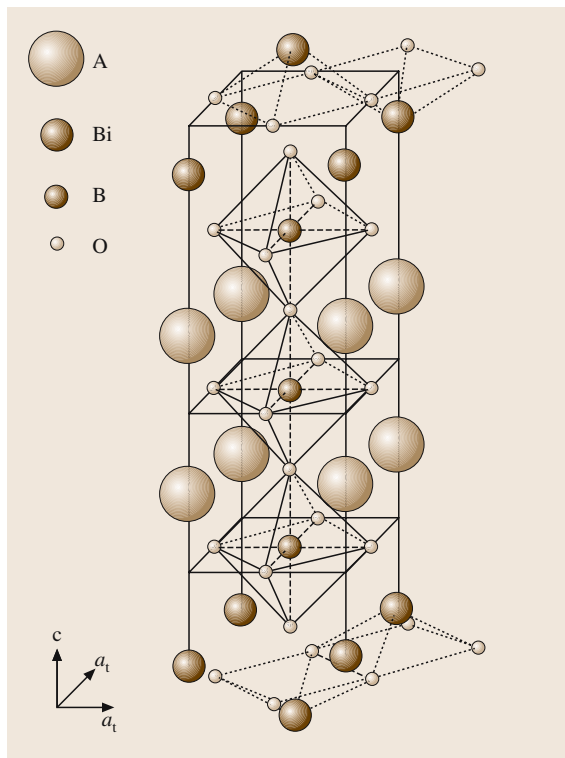


Fig. 27.11 Schematic diagram of half of the unit cell of the crystal structure of an Aurivillius compound with $m = 3$. The cell axes shown are those corresponding to the tetragonal structure

There are other examples of oxygen tetrahedral ferroelectrics. Lead germanate ($\text{Pb}_5\text{Ge}_3\text{O}_{11}$) is an interesting material which has been prepared as a single crystal. It is a hexagonal uniaxial ferroelectric which is also optically active. Ge can be substituted up to 50% by Si, which reduces the Curie temperature from 177°C to 60°C . The sense of the optical rotation switches with the sign of the spontaneous polarisation. This has been considered for optical devices and pyroelectric infra-red detectors, but has not reached any commercial applications. It has been prepared in ceramic and thin-film form and is occasionally used as a sintering aid in lead-containing ferroelectric ceramics. Gadolinium molybdate, $\text{Gd}_2(\text{MoO}_4)_3$ (GMO), possesses a structure consisting of corner-linked MoO_4 tetrahedra. It is an example of an improper ferroelectric material in which the phase transition from the paraelectric phase is controlled by the softening of a zone boundary mode, which would lead to a nonpolar low-temperature phase, which is in fact *ferroelastic*. The high-temperature phase is acentric, and the ferroelastic lattice distortion couples via the piezoelectric coefficients to give a spontaneous polarisation. The lattice distortion can be reversed by electrically switching the spontaneous polarisation and vice versa. GMO has no commercial applications, although it was at one time considered for optical switches.

27.1.2 Triglycine Sulphate (TGS)

The TGS family is a salt of the simplest amino acid, glycine ($\text{NH}_2\text{CH}_2\text{COOH}$) with sulphuric acid. In the crystal structure, the glycine groups are almost planar and pairs of them are connected by hydrogen bonds. A ferroelectric transition at 49.4°C between two monoclinic phases is driven by an ordering of the protons in these hydrogen bonds. The replacement of these protons by deuterons (which can be accomplished by repeated crystallisation of TGS from heavy water) to give DTGS results in a 10°C increase in T_C . There has been a great deal of academic interest in TGS because it possesses a well-behaved second-order order-disorder ferroelectric phase transition. In addition it has been of considerable technological interest for uncooled pyroelectric infra-red detectors and thermal imaging devices [27.20]. As a consequence, there has been considerable work on the replacement of various components with the objective of either improving performance or increasing T_C , such as can be achieved by deuteration. The replacement of glycine by l-alanine has been demonstrated to result in an internal bias field in the

crystal, which means that the crystals will retain their single-domain structure, even if heated well above T_C .

27.1.3 Polymeric Ferroelectrics

The first polymeric ferroelectric to be discovered was polyvinylidene fluoride (PVDF). The monomer for this possesses the formula CH_2CF_2 . It was first shown to be piezoelectric by Kawai [27.21] and later Bergman et al. [27.22] speculated about the possibility of it being ferroelectric, although no Curie temperature could be shown to exist, as the polymer melts first. The structure of the polymer is complex, as the polymer backbone can adopt a number of configurations depending upon whether the neighbouring carbon-carbon linkages adopt trans or gauche configurations. In a trans (T) configuration, the groups bonded to the carbon atoms sit on opposite sides of the carbon-carbon bond. In a gauche (G) configuration, they sit on the same side. In an all-trans configuration, the carbon backbone of the polymer forms a simple zigzag. In the case of PVDF, there are four different phases with different T and G sequences, and with different ways the polymers stack together in crystal structures. The polymer crystallises from the melt into *form II*, also called the α -phase, in which the bonds arrange themselves into the sequence TGT $\bar{\text{G}}$, see Fig. 27.12. The fluorine atoms are strongly electronegative which makes the C–F bond polar so

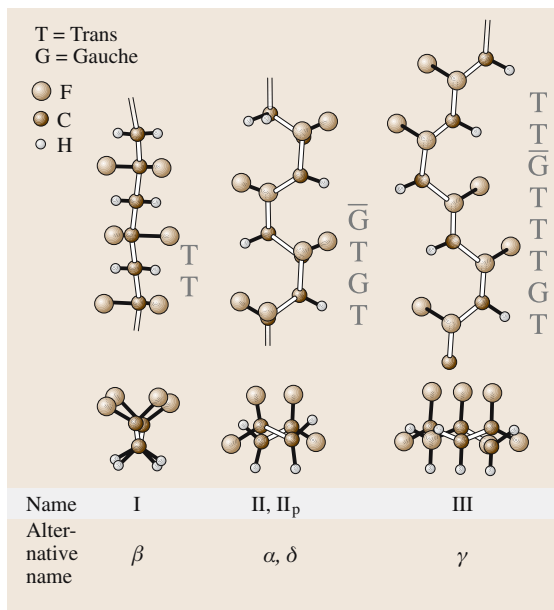


Fig. 27.12 PVDF bond structures

that the molecule has a net dipole moment perpendicular to its length. However, the molecules of polymer arrange themselves in the unit cell so that the dipoles cancel each other out. Form II is neither ferroelectric nor piezoelectric, but application of an electric field will convert it into form II_p, also called the δ -phase, in which the polymer molecules are arranged so that the unit cell has a net dipole moment. High-temperature annealing of either of these forms will produce form III (the γ -phase), which has a new TGT $\overline{\text{T}}$ TGTT configuration which also has a net dipole moment perpendicular to the long axis of the molecule and these arrange in a crystal structure which is also polar. Subjecting forms II or III to stretching or drawing will produce form I (the β -phase), which is an all-trans configuration (Fig. 27.12). It can be seen from Fig. 27.12 that of the three molecular configurations, the polar bonds all point most-nearly in the same direction and this is retained in the unit cell. Electrical poling makes this the most-strongly piezoelectric phase of PVDF. The stretching can be in a single direction, called uniaxial and usually achieved by drawing through rollers, or in two perpendicular directions, called biaxial and usually achieved

by inflating a tube of polymer. Strong electric fields are needed for poling and these are frequently applied by placing the polymer under a corona discharge. The formation of a copolymer of vinylidene fluoride with between 10% and 46% trifluoroethylene (TrFE) leads to a polymer which will crystallise directly into form I (the β -phase) from either melt or solution. This can be poled to produce a material which is as active as pure PVDF. The copolymers also show clear signs of a ferroelectric-to-paraelectric transition, such as exhibiting peaks in permittivity at T_C , with a T_C that depends on the amount of TrFE in the copolymer. Extrapolation to 0% TrFE implies a T_C in PVDF of 196 °C. PVDF is readily available and has achieved moderately widespread use as a piezoelectric sensor material. It is particularly useful where light weight and flexibility are important, or where very large areas or long lengths are needed. The P(VDF-TrFE) copolymers are not as readily available and are not widely used. Other polymers that have been shown to exhibit ferroelectric behaviour include the odd-numbered nylons, but these are only weakly piezoelectric and have not achieved any technological uses.

27.2 Ferroelectric Materials Fabrication Technology

Ferroelectrics are used commercially in a very wide range of forms, from single crystals through polycrystalline ceramics and thin films to polymers. Hence, only a summary of the fabrication techniques can be presented here.

27.2.1 Single Crystals

As has been shown above, ferroelectrics often have a wide range of complex compositions. This renders the problems of single-crystal growth much more difficult than for single-element crystals such as silicon. In many cases, the most technologically useful compositions are themselves solid solutions of complex ferroelectric end members. Frequently these do not melt congruently and sometimes one or more components of the melt will be volatile. Some of the growth techniques which are used are:

Czochralski Growth

Certain ferroelectric crystals melt congruently and can be pulled from the melt using the Czochralski technique. Examples include LiNbO₃ and LiTaO₃. Both materials are widely used technologically. LiNbO₃ is

used in surface acoustic wave (SAW) and electro-optic devices and LiTaO₃ in pyroelectric infrared detectors, piezoelectric resonators, SAW and electro-optic devices. LiNbO₃ melts congruently at 1240 °C, but the congruently melting composition is not at the stoichiometric composition (Li : Ta = 1), but lies at around 49% Li₂O. Crystals pulled from such melts in platinum crucibles are perfectly adequate for piezoelectric and most optical applications. However, it is known that the stoichiometric composition possesses rather better optical properties than the congruent composition and there is now a premium-grade material available commercially. LiTaO₃ melts congruently at 1650 °C. Again, the composition is not at Li : Ta = 1, but is slightly Ta deficient. The growth temperature is above the melting point for platinum. The solution is to use crucibles made from iridium, which is very expensive. Also, IrO₂ is volatile so the growth atmosphere must be N₂, which means that the crystals come out of the melt oxygen-deficient and must be embedded in LiTaO₃ powder and annealed under O₂ after growth to make them clear. Pt/Rh crucibles have been used to grow LiTaO₃ crystals, but the Rh absorbed into the crystals makes them brown. Such crystals are acceptable for piezoelectric and pyroelectric, but not

for optical, applications. Virtually all ferroelectric materials need to be electrically poled before use. In the case of LiTaO_3 , silver electrodes are applied to faces perpendicular to the polar (c) axis and a pulsed field of about 10 V/cm (1 ms with a 50/50 duty cycle) applied while the crystal is cooled slowly through T_C (620 °C). A similar process can be applied to pole LiNbO_3 crystals, or as the T_C is very close to the growth temperature, the field can be applied between the growing crystal and the crucible. This process has also been used to grow periodically poled crystals in which the field is reversed periodically during the growth of the crystal. Such crystals have some advantages in optical frequency-doubling in a process known as quasi-phase matching [27.23].

Solution and Flux Growth

The growth of crystals from aqueous solution is a technology very readily applied to water-soluble materials such as KDP and TGS. There are two processes that have been used. The first consists of rotating crystal seeds in a bath of saturated solution, which is slowly cooled so that material that comes out of solution is deposited on the seeds. The second consists of recirculating the bath solution over feed material, which is held at a slightly higher temperature and then back over the slightly cooler seeds, where the feed material is deposited. In both techniques, very careful control of temperature is essential (better than 0.002 °C) and the growth process can take many days.

The process of growing oxide single crystals from solution or flux is closely related to the above process, in that the crystals are grown from a saturated solution of the required oxide in a molten flux as it is slowly cooled. In the simplest version of this process, the oxides are sealed in a platinum crucible, heated to a *soak* temperature and slowly cooled to room temperature. Small crystals (usually up to a few mm in size) can then be recovered from the solidified melt. This has been applied to growth of a wide variety of small single crystals, such as PZT and PMN-PT solid solutions. PZT, for example, can be grown from $\text{PbO}-\text{B}_2\text{O}_3$ flux mixtures, or even pure PbO [27.25]. BaTiO_3 crystals can be grown from KF or TiO_2 fluxes. A major issue with the growth of Pb-containing ferroelectrics from PbO rich fluxes is the evaporation of PbO from the melt at the crystallisation temperature over the long time periods required to grow large crystals. This can be avoided to some degree by the use of sealed crucibles. A variation on the flux growth process is top-seeded solution growth, whereby a seed crystal is immersed from above into a supersaturated solution of the material to be grown. The crystal then grows

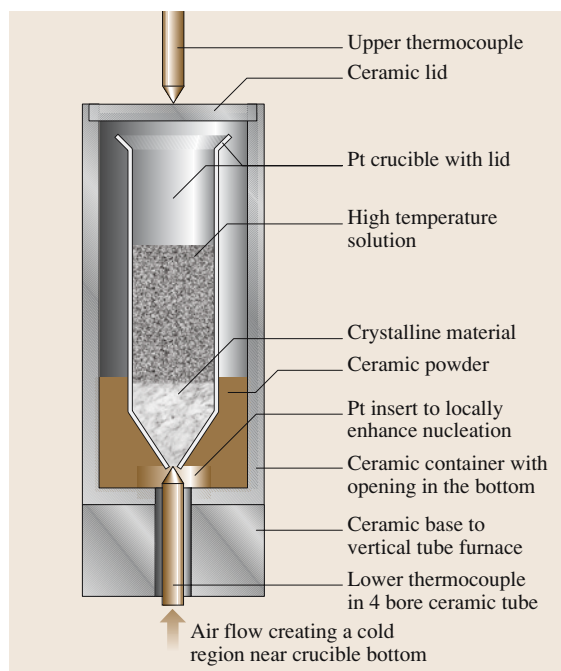


Fig. 27.13 A modified Bridgman apparatus used for growing piezoelectric single crystals (after [27.24])

as the solution is cooled further. The process is difficult to control and it has the major problem, especially when growing from a PbO flux, that the melt is open and therefore very prone to flux evaporation. It has been successfully applied to the growth of crystals of PMN-PT and PZN-PT solid solutions. The Bridgman process seeks to get over this problem by placing a dense compact of the flux/solute mixture inside a sealed crucible with a conical end, as shown in Fig. 27.13. A seed crystal is placed at the base of the cone which is held cooler than the rest of the crucible by a cooled rod. A molten zone is then allowed to travel up the crucible and the crystal solidifies inside it. Good-quality crystals have been grown this way and it is the preferred technique for the manufacture of PMN-PT and PZN-PT solid solutions. One problem with all flux-growth techniques is that the flux composition is constantly changing during the growth process and therefore the crystals tend to be nonuniform in composition. This is a particular problem for larger crystals, especially for solid solutions.

27.2.2 Ceramics

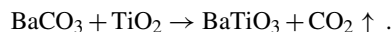
The vast majority of ferroelectric materials that are used commercially are used in the form of polycrystalline

ceramics. These are used in many of the applications listed above, including as dielectrics and in piezoelectric, pyroelectric, PTCR and electro-optic devices. Many different ferroelectric ceramic fabrication technologies have been developed over the years, but they are all based around a similar fabrication sequence. The key points are discussed below. Note that as the vast majority of commercially exploited ferroelectrics are oxides, the discussion will centre around these materials.

- **Raw-material selection:** the usual starting point for a ceramic fabrication sequence is the selection of the raw materials. For oxide ferroelectrics, the raw materials are usually oxides or carbonates, but occasionally other compounds (e.g. nitrates, citrates etc.) are used which will decompose when heated to form the oxides required. High purity (usually > 99.9% with respect to unwanted cations) is important for good reproducibility. Small quantities of dopants can have a major effect on the final electrical properties, as would be expected for any electronic material, but can also seriously affect the sintering characteristics and grain size of the final ceramic body. Note that high purity does not usually imply the kind of purity that would be needed for a semiconductor material. Indeed, ultra-high purity (> 99.99%) is often obtained by raw-material manufacturers by applying processes which seriously affect the reactivity of the oxide powders. A high degree of solid-state reactivity in the powders is important for the processes which follow. This is usually determined by a combination of raw-material particle size and specific surface area, although selecting the right crystallographic phase can be important. When using TiO₂, for example, it is usually found that the anatase phase is more reactive than rutile. These factors can be analysed for quality control purposes by using laser particle size analysis (taking care to break up loosely bound agglomerates by ultrasonic dispersion in water, with a dispersing agent prior to measurement), BET (Brunauer, Emmett and Teller)-specific gas absorption and X-ray powder diffraction respectively. The majority of ceramics are made by mixing together powdered raw materials, but occasionally solution mixing techniques are used and these will be discussed further below.
- **Raw-material mixing:** this is usually done in a ball mill consisting of a cylinder containing a mixture of small balls or cylinders made of steel, steatite, zircon or yttria-stabilised zirconia (YSZ). The oxide raw materials are accurately weighed into the ball mill,

together with a predetermined amount of a milling fluid (usually deionised water, but occasionally an organic fluid such as acetone). It is usual to add a dispersant to aid the breakup of agglomerates. The ball mill is then sealed and rotated to mix the ingredients. Precise milling conditions are determined by the materials being used, but the mixing time is usually of the order of a few hours. It is advantageous to use milling balls that are made of a material whose wear products will be reasonably innocuous in the final ceramic. For this reason, YSZ balls are preferred, as small amounts of iron and silica contaminants can have unwanted or deleterious effects on the properties of many ferroelectric ceramics. For similar reasons, the ball mills are frequently rubber lined. Many variations on this process have been explored. High-energy ball milling is receiving considerable attention. In this process, the milling balls are given very high energy by either aggressively vibrating the ball mill or by stirring them at high speed with a paddle. Raw materials trapped between the balls are both comminuted and, if the energy is sufficiently high, can be forced to react together. Crystalline raw materials can be made amorphous in this process. A significant amount of energy can be stored in the powders after this process so that subsequent sintering can be undertaken at lower temperatures. In an adaptation of this process, it can be done in a continuous flow-through mill, whereby the slurry is pumped through the high-energy mill, making it less of a batch process.

- **Drying:** the slurry that results from the mixing process is dried after the milling balls are removed. On a small scale, this can be done in an oven, but on a commercial scale, this is usually achieved by spray drying.
- **Calcination:** the purpose of this process is to react the raw materials into the required crystallographic phase. The dried powders are placed into a sealed crucible (usually high-purity alumina or zirconia), which is baked in a furnace at a temperature high enough to decompose any non-oxide precursors and cause a solid-state reaction between the raw materials, but not so high as to sinter the particles and form hard agglomerates that will be difficult to break up in subsequent processing. A simple example is the reaction between BaCO₃ and TiO₂ to form BaTiO₃:



This reaction will go to completion at about 600 °C. Note that a much higher temperature (> 1000 °C) is

required for the decomposition of BaCO_3 to BaO in the absence of TiO_2 .

- **Milling:** this is usually done in a similar manner to the mixing process described above, but is done for longer. The objective is to break up hard agglomerates and reduce the powder to its primary particle size. Again, high-energy milling techniques can be used. Dispersants are usually added to aid the process, and pH is controlled so that the powder does not flocculate. Organic binders can be added at this stage. The slurry from the milling process is usually spray-dried or freeze-dried as a free-flowing aggregated powder is required. The objective is to form soft aggregates that will break up on subsequent die pressing. Alternatively, the slurry (or *slip*) can be taken straight to a slip- or tape-casting process.
- **Shape forming:** after drying, the powders from the above process can be uniaxially pressed into the *green* shape required, typically using a steel punch and die-set. It is important in this process to make sure that the force is applied equally to top and bottom punches. Usually, the die is designed with a slight taper to ease removal of the workpiece after pressing. Faults that can occur at this stage of the process include *capping* or radial cracks in the work piece, which usually result from an uneven distribution of pressure. One way to avoid this is to use cold isostatic pressing. In this process the powder, or more frequently a lightly uniaxially cold-pressed green block, is vacuum-sealed into a rubber mould, which is then immersed in a bath of oil. The oil is taken up to a high pressure, which isostatically compresses the block. The advantage of using this process is that it avoids the nonuniform distribution of pressure that is frequently a problem with uniaxially cold pressing. The high pressure allows the agglomerates to be broken up and the process usually results in a much higher green density, which facilitates the sintering process. Another process that can be used to form the green item is slip-casting, which involves pouring the ceramic slip into a plaster mould, which is usually gently rotated. The plaster absorbs moisture from the slip and the ceramic particles are deposited as a layer. Surplus slip is poured off and the deposited layer can be separated from the mould after it is dry. This process has been used to make large cylinders of piezoelectric ceramics. A third process that is frequently used to make thin sheets of ceramic is tape-casting, by which the slip is passed under a set of *doctor blades* in the apparatus shown schemati-

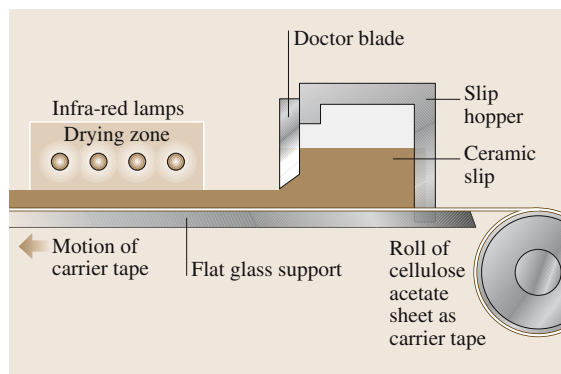


Fig. 27.14 Tape-casting apparatus

cally in Fig. 27.14. In this case, the slip is cast onto a continuous roll of plastic tape and a plastic binder is added to the slip which, when dry, holds the powder together and makes a flexible green tape. Sometimes the slip is cast directly onto a glass sheet. The green tapes can be screen-printed with patterns of metallic electrode inks and then laminated by warm-pressing to make multilayer structures with interleaved electrodes. This type of fabrication is widely used for multilayer ceramic (MLC) capacitors and multilayer ceramic piezoelectric actuators.

- **Sintering:** the green ceramic bodies are sintered in furnaces at temperatures of 1150–1300 °C, depending on the ceramic being manufactured. There is a good deal of process know-how in the sintering of the ceramic bodies. Key issues are:
 - The heating profile, which must be carefully controlled, especially over the lower range of temperatures up to about 500 °C where the organic matter (dispersants and binder) is burned off. Too rapid heating will cause large pores, or even cracks, to form.
 - For lead-containing ceramics, it is important that air be allowed free access to the vaporising organic material, or it will be carbonised and this can lead to reduction of the PbO in the ceramic to free Pb .
 - Loss of PbO by volatilisation at higher temperatures ($> \approx 800$ °C) means that the ceramic body needs to be sintered in a PbO -rich environment, which is often produced by packing the ceramic body in a *spacer powder* of PbZrO_3 , or **PZT** ceramic chips within well-sealed ceramic crucibles or saggars. This tends to restrict the access of air to the sintering ceramic body, which conflicts with the previous requirement.

- The ceramic body must be free to move during the sintering process, as there is significant linear shrinkage (about 15 to 18%). If the body is unable to slide over the surface on which it sits, this can lead to cracking, especially for large components. One solution to this problem is to sit the body on zirconia sand.

The sintering process has been very well described elsewhere. It is possible to obtain very high densities (> 98%) by careful control of the sintering conditions. For Pb-containing ferroelectric ceramics, a small excess of PbO is usually added to compensate for PbO loss by evaporation. This also tends to act as a liquid-phase-sintering aid, lubricating the grains of ceramic as they slide over each other in the sintering process, and providing a surface-tension force that pulls the ceramic grains together. The PbO also acts as a solvent, aiding the movement of the ceramic components during sintering and further densifying the ceramic. It is possible to sinter such materials to transparency, which means that there is virtually zero porosity. Very high densities can also be obtained by hot-pressing, in which the green body is placed inside a ceramic punch and die-set and raised to the sintering temperature under a pressure of about 35 MPa. Alternatively, hot isostatic pressing (HIP) can be used. In this process the pressure transmitting medium is a high-pressure gas (usually argon, but if the ceramic is Pb-containing it is important to include a few percent of oxygen to prevent reduction of the ceramic). In this case pressures of 100 MPa or more can be used. If the ceramic is pre-sintered so that there is no open porosity, then this can be achieved without the need for any container, but if there is open porosity, the body must be encapsulated in a suitable metal container, which will usually need to be a noble metal such as Pt.

- Electroding: good-quality electroding is essential for all ferroelectric devices for a variety of reasons. Any low-permittivity layer between the electrode and the ferroelectric material will manifest itself in a fall-off in capacitance and an increase in loss as the frequency of measurement is increased. Poor-quality electrodes, or even the wrong type of conductive material can lead to problems with device ageing, or even an inability for the device to function as intended. Most piezoelectric materials are supplied with a *fired-on* silver electrode, which is a mixture of silver powder, a finely divided glass frit and a fluxing agent. Obtaining the correct balance between metal and glass contents is important, as too high

a proportion of metal will lead to poor electrode adhesion while too high a proportion of glass will lead to poor electrode conductivity. Such electrodes are solderable with the use of appropriate fluxes. Sputtered or evaporated metal electrodes such as Ni or Cr/Au can also be used, as can metal electrodes deposited by electroless processes (e.g. Ni). The ohmic nature of the electrode contact to the ceramic is not usually important for highly insulating materials, such as piezoelectrics, but is important for semiconducting ceramics such as PTCR BaTiO₃. In this case it is usual to use Ni electrodes, which need to be annealed after deposition to develop the ohmic contact. In some cases it is necessary to fire on the electrode at the same time as ceramic sintering (cofired electrodes). This is particularly important where the electrodes are buried in the structure, as with MLC capacitors and actuators. Clearly, there are problems to solve here in terms of the potential oxidation or melting of the electrode material. The conventional materials to use in such electrodes have been noble metals such as palladium (sometimes alloyed with silver) or platinum, which are very expensive. There have been serious efforts to develop ferroelectric compositions that can be fired with base metal electrodes such as Ni. In the case of BaTiO₃-based dielectrics, this has entailed the development of heavily acceptor-doped compositions that can stand being fired in a neutral or slightly reducing atmosphere. In the case of MLC actuators, there has recently been some success in developing piezoelectric PZT compositions that can stand being fired in such atmospheres, using Cu as the electrode material.

- Poling: many devices made from ferroelectric ceramics (all piezoelectric and pyroelectric devices) require poling before they will develop useful properties. This entails applying a field that is significantly in excess of the coercive field (typically 3×), usually at an elevated temperature. It is not usually necessary to exceed the Curie temperature. For example, PZT ceramics with T_C in the range 230–350 °C can be poled by applying 35–kV/cm (depending on the type of PZT – see below – soft PZTs need lower poling fields than hard PZTs) at about 150 °C, with the field kept applied while the workpiece is cooled to room temperature. It is usual to immerse the ceramic in a heated bath of oil (mineral or silicone) during the process. One disadvantage of this is that the ceramic then needs to be carefully cleaned after poling. Silicone oil can

be very hard to remove completely and its presence as a residue will compromise electrode solderability. For this reason, some workers have developed a process whereby the ceramic is poled under SF_6 gas. There is a rapid decay of properties after poling. This decay stabilises after a few hours, so it usual to wait at least 24 hours before electrical properties are measured.

There are many variations on the above basic process route which have been researched. One of these is the use of solution techniques to prepare the oxide powders. The basic principle here is that if the cations are mixed in solution, then they will be mixed on the interatomic scale without the need for milling processes that can introduce impurities. (This is a matter of discussion, as the act of precipitation and decomposition can lead to separation of the components.) Many routes have been explored, including the use of inorganic precursors such as nitrates and metalorganic precursors such as oxalates, citrates or alkoxides and acetates. Some of these have achieved a degree of commercial success, although the use of solution routes is more complex and the raw materials much more expensive than the mixed oxide routes. The use of mixed barium titanium oxalates has been very successful in producing high-quality fine-grained barium titanate powder for use in the capacitor industry. The use of metal citrates (frequently called the Pechini process) has been successfully used on a research scale to prepare many different types of ferroelectric oxide, but this type of process has not been applied on a commercial scale. Metal alkoxides, such as titanium isopropoxide, are readily soluble in alcohols and will react quickly with water to precipitate a hydroxide gel. Workers have used mixtures of titanium and zirconium alkoxides with lead and lanthanum acetates to coprecipitate a mixed hydroxide gel that could be calcined and sintered to make transparent lead lanthanum zirconate titanate ceramics for electro-optical applications.

27.2.3 Thick Films

There has been considerable interest in the integration of thick (10–50 μm thick) films of ferroelectric materials with alumina and other types of substrates such as silicon, to complement the wide range of other thick-film processes that are available, covering conductors, dielectrics, magnetic materials etc. There are many potential advantages to thick film processing for making certain types of sensor, especially the ability to use screen-printing for depositing the patterns of the ma-

terials required. Screen-printing involves using a sheet of wire mesh (the *screen*) that is coated with a photosensitive polymer. Exposure and development of the polymer allows selected areas to be removed, opening regions through which a paste of the required material can be pushed using a rubber blade or squeegee. The principle is simple, but there is a considerable amount of know-how in the formulation of the paste, which consists of the active material (in this case a ferroelectric powder such as PZT), an organic vehicle (a mixture of a solvent and polymer) and a glass frit. The screen is stretched over a former, and held close to, but not in contact with, the surface onto which the print is required. The paste is placed on the screen, and then spread over the screen with the squeegee, which prints the paste onto the substrate. Successive layers of different materials can be printed and cofired, provided there is good compatibility between them. The process has been well developed for piezoelectric films and adequate properties have been obtained from the films, although they are still well below the values that could be expected from a bulk ceramic material. (See review by Dorey and Whatmore [27.26] for further details.)

27.2.4 Thin Films

The integration of high-quality thin films ($< 0.1\text{--}5\ \mu\text{m}$ thick) of ferroelectric materials onto substrates such as silicon has excited considerable interest for potential applications ranging from nonvolatile information storage to their use as active materials in micro-electro-mechanical systems (MEMS), where they can potentially be used for microsensors and actuators. Almost all the interest has been in the use of oxide ferroelectrics, but there has been some interest in the use of P(VDF-TrFE) copolymer films. These can be spun onto electroded substrates from methyl ethyl ketone solution. They are dried at relatively low temperatures ($< 100\ ^\circ\text{C}$) and crystallised by annealing at $180\ ^\circ\text{C}$ for several hours [27.27]. Such films have been applied to pyroelectric devices. However, the activity coefficients which can be obtained from such films are much lower than those that can be obtained from oxide materials. A range of deposition techniques have been developed for growing ferroelectric oxide films, which are summarised below:

- Chemical solution deposition (CSD): this term is applied to a wide range of processes that involve taking the metal ions into metalorganic solution, which is then deposited on the substrate by spinning, fol-

lowed by drying and annealing processes to remove the volatile and organic components and convert the layer into a crystalline oxide. There are two broad classes of CSD process: metalorganic deposition (MOD) and sol gel. The MOD processes usually involve dissolving metal complexes with long-chain carboxylic acids in relatively heavy solvents such as toluene. The carbon content of the precursors is quite high, so there is a good deal of thickness shrinkage during firing. MOD solutions tend to be quite stable with time and resistant to hydrolysis. Sol-gel processes use precursors such as metal alkoxides, acetates and β -diketonates in alcohol solution. (For example, a set of precursors to deposit PZT would be Ti isopropoxide, Zr n-propoxide and lead acetate). Alkoxide precursors are very susceptible to hydrolysis, and so careful control of moisture content during the sol synthesis is essential and stabilisers such as ethylene glycol are usually added to the solutions to extend the useable lifetimes of the sols. Whereas MOD solutions are true solutions, the sols are actually stable dispersions of metal oxide/organic ligand particles with a size of about 4–6 nm. Sols possess lower viscosities and tend to produce the oxide layer at a somewhat lower temperature than the MOD processes, but the individual layer thicknesses produced by a single spin tend to be lower. Single crack-free layers tend to be in the range 100–200 nm thick. CSD processes have the advantages that they are low cost, the composition can be easily changed and they produce very smooth layers. The processes are planarising and will not follow underlying surface topology, which can be a disadvantage. Also, the processes are not industry standard in that they are wet and tend to have many variations.

- Metalorganic chemical vapour deposition (MOCVD): this is a variation on the process that has been very successfully applied to the growth of group III–V semiconductor layers. The principle is simple: volatile metalorganic compounds are passed over a heated substrate, where they decompose to form a layer of the desired compound. The problem with the growth of ferroelectric oxides is that most of the available metal organic precursors are relatively nonvolatile at room temperature. There has, therefore, been a great deal of research into the available precursors for the compounds that are required. Metal alkyls (such as lead tetraethyl) are very volatile, but only available for relatively few of the metal ions of interest (Pb being the main one). They are also pyrophoric and highly toxic.
- Some metal alkoxides, such as Ti isopropoxide, are suitable MOCVD precursors. Metal β -diketonates and related compounds such as tetramethyl heptanedionates (THDs) have received considerable attention as Ba and Sr precursors. All of these precursors need to be heated to give them suitable volatility and this means that the lines connecting the precursor source to the growth chamber need to be heated as well. Some of the precursors (especially THDs) are solids, which means that they are not really suitable for use in conventional bubbler-type sources. There has been considerable success in using solutions of these compounds in tetrahydrofuran (THF). The solutions are sprayed into a vaporiser that consists of a cylinder, containing wire wool or ball bearings, heated to a temperature at which the solution will flash-evaporate. A carrier gas is passed through the cylinder and this carries the precursor vapour into the growth chamber. The growth chamber is usually held at reduced pressure and a certain amount of oxygen is introduced to aid the oxide deposition. Frequently a radio-frequency (RF) or microwave plasma is also introduced to aid the growth of a high density film and reduce the required substrate temperature. The major problem with the MOCVD process for complex ferroelectric oxides which have many cation components is finding the right combination of precursors that will all decompose at the same substrate temperature (usually ca. 550–650 °C) at a rate that will give the desired composition in the film. The process has been very successful in growing thin films of materials such as (Ba, Sr)TiO₃, with potential applications in the dynamic random-access memory (DRAM). The major advantage of MOCVD is that it is a truly conformal growth technique, with major advantages for semiconductor devices with complex topologies, but is a very expensive technique to set up because of the complex growth and control systems needed. Also, precursor availability is still a problem for many systems.
- Sputtering: a range of sputtering processes have been applied to the growth of ferroelectric thin films, including RF magnetron sputtering, direct-current (DC) sputtering and dual ion-beam sputtering. The RF magnetron process is probably the most popular. With all the processes, the major problem is one of obtaining the correct balance of cations in the growing film. Many different solutions have been found to this problem. In reactive sputtering, a composite metal target can be used. This can be made of segments of the metals to be sputtered (for exam-

ple, Pb, Zr and Ti for PZT) and their relative areas changed to obtain the right composition in the film. Alternatively, multiple targets can be used and the substrate exposed to each one for different lengths of time, or the power applied to each one can be varied. In reactive sputtering, it is necessary to have an amount of oxygen in the sputtering gas (usually Ar). It is possible to sputter ferroelectric thin films from ceramic or mixed powder targets, but it is necessary to adjust the target composition to allow for different yields for each element. In any sputtering process there are many variables to adjust to optimise the process, including the sputtering power, and RF or DC substrate bias, which will affect the ion bombardment of the growing film, the sputtering atmosphere pressure and gas mixture and the substrate temperature. All of these can affect the film growth rate, composition, crystallite size and crystalline phases that are deposited, and the stress in the growing film. For this reason, the development of a sputtering process for a complex ferroelectric oxide can be a time-consuming business, and once a set of conditions has been arrived at for one particular composition, it cannot quickly be changed to accommodate new compositions. Dual ion-beam sputtering differs from the RF and DC processes in that a much lower background pressure is used, the material is sputtered from the target using an ion beam and a second lower-energy ion beam is used to stimulate and densify the growing film. The sputtering processes have the advantage of being well accepted industrially, as they are dry and can readily coat large-area substrates.

- Laser ablation: this process involves bombarding a ceramic target with a pulsed, focussed laser beam, usually from an ArF excimer source. The target is held under vacuum. A plasma plume is produced and the products ablated from the target are allowed to fall on a heated substrate. The advantages of the process are that there is usually good correspondence between the composition of the target and the growing film. Relatively small ceramic targets are acceptable for the process, and it is thus a good method for getting a rapid assessment of the properties of thin films of a given material. The disadvantages of the process are that the plasma plume will only coat a relatively small area of substrate, although there are now systems which use substrate translation to coat large areas, and particles can be ablated from the target, causing defects in the growing film.

In all the techniques used for the growth of ferroelectric oxide thin films, the key issues are control of composition and the formation of the desired crystalline phase (usually perovskite) with the desired crystallinity (crystallite size, morphology and orientation). All of the ferroelectric perovskites have a tendency to crystallise into a non-ferroelectric fluorite-like pyrochlore phase at low temperatures. In the case of the CSD processes, this means that as the film is heated from room temperature, after it loses the organic components, it first forms an amorphous oxide which then crystallises into a nanocrystalline pyrochlore phase, finally forming the desired perovskite phase. The temperatures at which this will occur depend very much on the ferroelectric oxide that is being grown and the precise composition. In the case of PZT, the pyrochlore phase will form above about 300–350 °C. The perovskite phase will start to form above about 420 °C, depending upon the ratio of Zr : Ti in the solid solution. The compositions close to PbTiO₃ will crystallise into perovskite much more readily than those close to PbZrO₃. In the case of a complex perovskite, such as Pb(Mg_{1/3}Nb_{2/3})O₃ or PbSc_{1/2}Ta_{1/2}O₃, the pyrochlore phase is much more stable and much higher temperatures (> 550 °C) are needed to convert it to perovskite. Excesses of PbO will tend to favour the formation of perovskite, and deficiencies favour pyrochlore. Higher annealing temperatures will promote PbO loss and it is possible to get into a position, through PbO loss, where the pyrochlore becomes the most stable phase, even at high annealing temperatures. Residual pyrochlore phase invariably compromises the electrical properties of the films through reduced permittivity and piezoelectric/pyroelectric coefficients. The other growth techniques have the advantage that the films can be deposited onto heated substrates, at temperatures where they will grow directly into the perovskite phase (at least in principle), although there are many examples in the literature of films being deposited (e.g. by sputtering) at low substrate temperatures and converted to the desired perovskite phase by post-deposition annealing, in which case the same problems of pyrochlore formation apply. The control of film crystallinity (crystallite orientation and size) is important as it has a direct effect on the electrical properties. This is usually achieved through control of the crystallite nucleation. Sputtered Pt is frequently used as a substrate onto which ferroelectric thin films are grown. Like many metals, this will naturally grow with a (111) preferred orientation. It is face-centred cubic (FCC), with a lattice parameter of about 3.92 Å, which matches the lattice parameters of many of the ferroelectric perovskites, which are about

4 Å. This means that, with appropriate process control, it is quite possible to get a highly orientated (111) ferroelectric film on Pt, with a crystallite size of about 100 nm. Changing the underlying nucleation layer can allow other orientations to be grown. For example, the use of thin films of TiO₂ or PbO on top of the Pt electrode can induce a (100) orientation. It is important to

control the nucleation density of the perovskite phase. If this is allowed to become too low, than large circular grains several microns in diameter (called *rosettes*) can form, which tend to induce defects at their boundaries. (Further details on thin-film ferroelectric growth techniques can be found in the book by *Pas de Araujo* et al. [27.28]).

27.3 Ferroelectric Applications

27.3.1 Dielectrics

As noted above, BaTiO₃ possesses a very high permittivity close to T_C . The inclusion of selected dopants can reduce T_C and optimise the properties of the material for capacitor applications. The substitution of Sr²⁺ for Ba²⁺, for example, will reduce it, so that at about 15% substitution it will occur at about 20 °C. Substitution of Zr⁴⁺ for Ti⁴⁺ has a similar effect. The use of off-valent substitutions will have an effect upon resistivity and degradation characteristics, so that substituting La³⁺ on the A site will reduce resistivity at low concentrations. Substituting Nb⁵⁺ on the B site at the 5% level has been shown to confer resistance to degradation. The addition of Mn to the lattice has been shown to have a positive effect on dielectric loss. Control of grain size has also been shown to have a marked effect on dielectric properties. Reducing the grain size has the effect of increasing the concentration of domain walls per unit volume of ceramic, and thus increasing the domain-wall contribution to the permittivity. A reduction in grain size also brings about an increase in the unrelieved stress on the grains, which further increases the permittivity. Once the grain size falls below about 0.5 µm, the stress on the grains reduces their tetragonality and the permittivity falls. It is also possible to use heterogeneity in the ceramic grains to flatten the curve of dielectric constant versus temperature. A range of Electronic Industries Alliance (EIA) codes have been introduced to define the variation of capacitance with temperature. The X7 temperature range, for example, is −55–+125 °C, while following this with the letter R would specify a capacitance change of no more than ±15%. Other dielectrics have been developed based on PMN that have higher peak permittivities, but worse temperature characteristics, capable of meeting a Z5U specification (+22 to −56% capacitance variation over the range 10–85 °C) but not much better. There has been huge progress in MLC technology, with capacitors now available with

> 100 layers of submicron thickness and capacitances ranging from a few hundred pF to 100 µF. The majority of these now have base-metal (Ni) electrodes. The fabrication of the ceramic dielectrics for these is a very complex area and a great deal of technology has been developed to service a very large MLC capacitor market (615 Bn units in 2002, worth an estimated \$ 8.4 Bn).

27.3.2 Computer Memories

The storage of digital information is of great interest and growing technological importance. The ability of ferroelectrics to store information via the sense of the spontaneous polarisation has made them candidates for this application ever since the advent of the electronic computer. However, it is only with the development of the techniques for the growth of ferroelectric thin films onto silicon at relatively low temperatures (see above) that their use has become a reality. There are two ways in which ferroelectrics can be used in computer memories. The first is the replacement of the dielectric layer in dynamic random-access memories (DRAMs). Here, the motivation is to exploit the high permittivities exhibited by ferroelectric oxides to reduce the area of silicon required to store a single bit of information. The favoured materials here are based on (Ba, Sr)TiO₃, while the favoured technique for thin-film growth is plasma-enhanced MOCVD because of its ability to deposit conformal coatings. No DRAM devices using these thin films have yet reached the market place. The use of the switchable polarisation for the nonvolatile storage of information has received a great deal of research. A good deal of this work has been based upon the use of PZT thin films, grown by a variety of techniques, including CSD, RF sputtering and MOCVD. Initially, there were many problems to solve, including the decay of the switchable polarisation (ca. 35 µC/cm²) as the number of switching cycles increased (fatigue), the tendency for the polarisation in a bit to become less easily switched with time

and the number of times it is switched in the same sense (imprint) and the tendency for the switchable polarisation in any one bit to decay with time (retention). The fatigue and imprint problems in **PZT** have been ascribed to the motion of oxygen vacancies addressed by the use of oxide, rather than metallic, electrodes and doping to reduce the vacancy concentration. It has also been demonstrated that aurivillius compounds such as strontium bismuth tantalate (SBT) and La-doped $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ not show fatigue behaviour, although they do require rather higher growth temperatures than the **PZT** family and have a significantly smaller switchable polarisation (ca. $7 \mu\text{C}/\text{cm}^2$). A number of companies are now making chips with nonvolatile on-board embedded ferroelectric random-access memory (FRAM). These are going into applications such as smart cards and computer games.

27.3.3 Piezoelectrics

This represents the most diverse range of applications for ferroelectric materials, as noted in the introductory section, covering a wide range of sensors, actuators and acoustic wave components for frequency control and filtering applications. The key properties determining the performance of a piezoelectric material in a particular application are the piezoelectric coefficients, elastic and dielectric constants. These are all tensor properties (see Nye [27.29]) and are combined through the constitutive equations:

$$D_i = d_{ijk} X_{jk} + \varepsilon_{il} E_l,$$

$$x_{ij} = s_{ijkl}^E X_{kl} + d_{mij} E_m,$$

where D_i = electric displacement, E_i = electric field, X_{ij} = stress, x_{ij} = strain, d_{ijk} = piezoelectric coefficients, ε_{ij}^X = dielectric permittivity coefficients at constant stress, and s_{ijkl}^E = elastic compliance coefficients at constant electric field.

The above equations use the Einstein repeated-suffix tensor notation. Note that most piezoelectric equations use the reduced-suffix notation for the piezoelectric and elastic constants [27.29].

There are two basic modes of operation for piezoelectric sensors and actuators. The first is well away from any mechanical resonance – usually at low frequency or quasi-DC. Here, the electric signals generated by an applied stress, or the strain generated by an applied electric field can be easily calculated using the constitutive equations of the type quoted above. The commonly used modes for operating piezoelectric ceramic actuators are longitudinal, where the electric field is applied

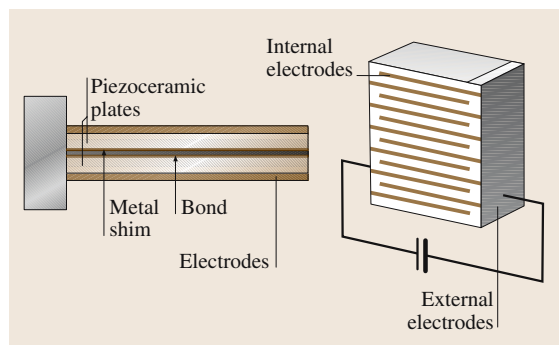


Fig. 27.15 (a) Piezoelectric bimorph and (b) MLC actuator structures

parallel to the direction in which strain is required and the coupling is via the piezoelectric d_{33} coefficient, and transverse, where a field is applied along the polar axis, but the exploited strain is perpendicular to this direction and coupling is via the d_{31} and d_{32} coefficients. Corresponding modes are frequently used for piezoelectric sensors. The two most commonly used actuator structures are illustrated in Fig. 27.15. Figure 27.15a shows a piezoelectric bimorph, in which two pieces of piezoelectric ceramic are bonded to a central metal shim. Fields are applied parallel to the polar axes of the ceramic elements so that their transverse extensions are equal and opposite. Usually one end of the bimorph is clamped so that the other end bends. In a variation on this structure, a single piece of ceramic is bonded to a piece of metal, making a unimorph. Bimorph and unimorph devices can provide high displacements (several tens to thousands of microns at up to 200-V drive) but relatively low forces (typically up to 1 N), depending on the size of the bimorph. The second widely used actuator structure is the multilayer ceramic actuator (Fig. 27.15b). This device operates in longitudinal mode, with applied field and extension parallel to the polar axis. The division of the structure into many thin layers reduces the voltage required to obtain a given extension. Typical devices can produce a few to a few tens of microns extension at a few hundred Volts (depending on dimensions) but high forces (in the kiloNewtons range). It is also possible to use a shear mode of operation for both sensing and actuation. In this actuation mode, the field is applied perpendicular to the polar axis and a shear strain is generated in the plane containing the vectors parallel to the applied field and the polar axis.

Many piezoelectric devices are operated under an alternating electric field whose frequency is tuned to match a mechanical resonance. A piezoelectric element oper-

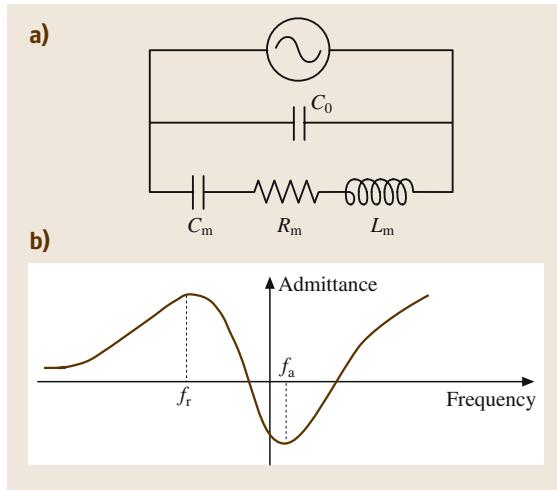


Fig. 27.16 (a) Piezoelectric element equivalent circuit (b) Admittance versus frequency for a typical piezoelectric resonator

ated well away from resonance will exhibit a capacitive impedance characteristic. As a mechanical resonance is approached, the electrical behaviour of the element can be characterised using an equivalent circuit as shown in Fig. 27.16a. C_0 is the static capacitance. The L_m , R_m , and C_m form the motional arm of the circuit. If the admittance of such a circuit is plotted as a function of frequency, a behaviour is obtained as shown in Fig. 27.16b. The maximum in the admittance is very close to the mechanical resonant frequency (f_r), while the minimum in the admittance is close to antiresonance (f_a); f_r is determined by the elastic properties of the piezoelectric, its density and the dimensions relevant to the particular resonant mode. It is useful to define an effective electromechanical coupling coefficient k_{eff} which is defined by:

$$k_{\text{eff}}^2 = \frac{\text{mechanical energy converted to electrical energy}}{\text{input mechanical energy}}$$

or

$$k_{\text{eff}}^2 = \frac{\text{electrical energy converted to mechanical energy}}{\text{input electrical energy}}$$

k_{eff} can be determined from f_r and f_a :

$$k_{\text{eff}}^2 \approx \frac{f_a^2 - f_r^2}{f_a^2}$$

It can be shown that, for a simple resonant mode of a piezoelectric element, k_{eff} takes the form:

$$k_{\text{eff}}^2 = \frac{d^2}{s^E \epsilon^X}$$

where the d , s and ϵ coefficients in this equation are combinations of the piezoelectric, elastic and dielectric permittivity coefficients. The precise combination of the coefficients will depend on the shape of the piezoelectric element and the mode being excited into resonance. A list of the more commonly used modes and the dependence of k_{eff} on the relevant materials constants is given in Table 27.4. For more complex structures, such as those which involve a piezoelectric element (or elements) bonded to a metal component, it is still possible to define and measure a k_{eff} , but its value will be a much more complex combination of the materials constants, and will also involve the properties of the metal parts and the relative dimensions of the piezoelectric and metal components. (For more details refer to the IEEE standards on piezoelectricity).

By far the most widely exploited piezoelectric materials system is based on the $\text{PbZr}_x\text{Ti}_{1-x}\text{O}_3$ ceramic solid solution series. Most of the compositions of-interest are centred on the morphotropic phase boundary (MPB) region (Fig. 27.7) with $x = 0.52$. In the undoped series, the piezoelectric coefficients and dielectric constants peak at this composition. However, their properties can be greatly enhanced by the inclusion of selected cation dopants. These can be classified as follows:

- **Isovalent:** these are divalent cations that will substitute for Pb^{2+} on the A site (e.g. Sr^{2+}) or tetravalent ions (e.g. Sn^{4+}) substituting for Ti/Zr^{4+} on the B site.
- **Donors:** these are cations with a valency greater than the site onto which they substitute (e.g. La^{3+} for Pb^{2+} or Nb^{5+} for Ti/Zr^{4+}).
- **Acceptors:** these are cations with a valency lower than the site onto which they substitute (e.g. K^+ for Pb^{2+} or Ni^{3+} for Ti/Zr^{4+}).
- **Multivalent:** these are ions that take multiple valency (e.g. Mn substituted on the B site).

The broad effects of each class of dopants are listed in Table 27.5.

The resulting **PZT** ceramics are broadly classified into *hard* or *soft* according to whether they are doped with acceptors or donors, respectively. Each manufacturer has its own nomenclature scheme for its piezoceramic products. However, the US Navy produced a set of specifications under MIL-STD-1376B

Table 27.4 A list of the most commonly used piezoelectric resonances and the dependence of the k_{eff} for each on the piezoelectric, elastic and dielectric constants [27.30]

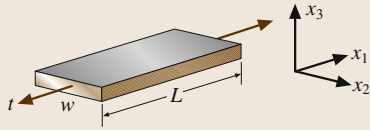
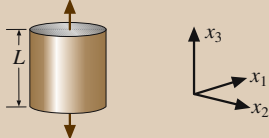
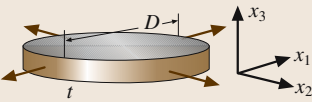
Mode type	Mode shape	Coupling factor
Length extensional mode of a bar using transverse excitation $L > 5w, L > 10t$		$k_{31}^l = d_{31} / \sqrt{\epsilon_{33}^T S_{11}^E}$
Length extensional mode of a cylindrical rod using longitudinal excitation $L > 10 \times \text{diameter}$		$k_{33}^l = d_{33} / \sqrt{\epsilon_{33}^T S_{33}^E}$
Radial mode of a round plate using longitudinal excitation $D > 20 \times \text{thickness}$		$k_p = k_{31} / \sqrt{2/(1-\sigma^p)}$
Grey areas denote the electroded surfaces		

Table 27.5 The effects of substituents in the PZT system (after [27.8])

Isovalent substituents e.g. Sr^{2+} for Pb^{2+} , Sn^{4+} for Zr^{4+}	Donor substituents e.g. La^{3+} for Pb^{2+} , Nb^{5+} for Zr/Ti^{4+}
<ul style="list-style-type: none"> Lower Curie point Increase permittivity Small improvement in linearity at high drive No change in coupling factor, aging, volume resistivity or low amplitude mechanical or dielectric loss 	<ul style="list-style-type: none"> No change in Curie point Higher permittivity Higher electromechanical coupling Poorer high drive linearity
	<ul style="list-style-type: none"> Lower aging Increase resistivity Reduced Q_E and Q_M
Acceptor substituents e.g. K^+ for Pb^{2+} ; Fe^{3+} for Zr/Ti^{4+}	Variable valence substituents e.g. Mn – can be 2+, 3+, 4+ – for Zr^{4+}
<ul style="list-style-type: none"> No change in Curie point Lower permittivity Lower electromechanical coupling Improved drive linearity Little change in aging Volume resistivity reduced Increased Q_E and Q_M 	<ul style="list-style-type: none"> No change in Curie point Range of permittivity Range of electromechanical coupling – usually lower High drive/linearity not changed Reduced aging

which are universally recognised. Hard ceramics (e.g. US Navy type III) have relative permittivities in the range of 1000 with high Q_s (> 500) and d_{33} of about 215 pC/N. Soft ceramics (e.g. US Navy type VI) have much higher permittivities (3250) and piezoelectric coefficients ($d_{33} = 575$ pC/N) but lower Q 's (65). Hard ceramics have much better drive-voltage stability and higher Curie temperatures than soft ceramics. The soft compositions tend to be

used for actuator applications, where the maximum piezoelectric displacement for a given drive voltage is sought, while soft ceramics tend to be used for resonant and sound-generation applications. There are, however, exceptions to this. Medical ultrasound transducer arrays tend to use soft, high-permittivity ceramics, because they need to have the maximum charge sensitivity in the reception of the sound waves reflected from tissues within the body, and

Table 27.6 Typical values for the piezoelectric, dielectric and elastic properties of some selected piezoelectric ceramic materials (taken from [27.31])

Material type	I	III	V	VI	Units
Stress-free relative permittivity (ϵ_{33})	1275	1025	2500	3250	
Dielectric loss tangent ($\tan \delta$)	≤ 0.006	≤ 0.004	≤ 0.025	≤ 0.025	
d_{33}	290	215	495	575	pC/N
k_p	0.58	0.50	0.63	0.64	
Mechanical quality factor Q_m	≥ 500	≥ 800	≥ 70	≥ 65	
Curie temperature (T_C)	325	325	240	180	°C

high permittivities because the small elements usually need to drive transmission lines with significant capacitance.

There are other types of piezoelectric ceramics that are used for specialist applications. Modified PbTiO₃ ceramics possess very low values of d_{31} and tend to be used for ultrasound array applications where small cross-coupling between adjacent elements is important. They also possess much higher response to hydrostatic stress than MPB PZT ceramics. Because of this, arrays of small blocks of this type of ceramic are used in flank array sensors in submarines. As noted above, lead metaniobate ceramics are used in high-temperature applications. Table 27.6 lists the piezoelectric and other properties of some selected piezoelectric ceramics.

Single crystals of PZN-PT and PMN-PT have received considerable attention because of the very high piezoelectric d_{33} coefficients and k_{33} coupling factors that can be obtained and single-crystal plates of these are now commercially available. Table 27.7 lists some of the properties that have been measured from these materials, and compares them with a commercial soft PZT ceramic.

Single crystals of LiNbO₃ are widely used in surface acoustic wave devices. In these, interdigitated electrodes (IDE) are applied to one polished face of a crystal. Excitation of these with an RF field will cause the generation of piezoelectrically excited Rayleigh waves, which will propagate with very little attenuation. These can be detected by further sets of IDE placed on the same face of the crystal. Such devices are widely used in signal processing and filtering applications in mobile telecommunication applications.

Thin films of piezoelectric materials, excited into thickness-mode resonance, are being explored for very-high-frequency (> 1 GHz) resonant filter applications.

27.3.4 Pyroelectrics

The pyroelectric effect is now widely used in uncooled detectors of long-wavelength infrared radiation (IR). The principle behind the operation of these devices is very simple. The radiation to be detected is allowed to fall upon a thin chip of the pyroelectric material. The energy absorbed causes a change in temperature and the generation of a pyroelectric charge, which will cause the flow of current in an external circuit. This can be amplified and used, for example, to switch an alarm. The basic circuit of a pyroelectric detector is shown schematically in Fig. 27.17. The field-effect transistor (FET) amplifier needs to be close to the pyroelectric element, ideally in the same package. This is because the latter has a very high output impedance and thus is very noise sensitive. The pyroelectric signal is represented in this circuit as a current source, i_p . i_p is proportional to the rate of change of the element temperature with time. Pyroelectric devices do not have a DC response. They only see changes in the intensity of the radiation with time. This is a major advantage in many applications, where usually there is a requirement to detect changes in the radiation coming from a scene, rather than the average or unchanging background intensity. An example of this is the requirement to detect the radiation from a person coming into the field of view of an IR detector.

Pyroelectric devices have several advantages over other radiation detectors. Their response is independent of the wavelength of the incident radiation, provided there is some means to absorb the radiation. Hence, pyroelectric radiation detectors have been used across the full range of the electromagnetic spectrum, from microwaves to X-rays. Basically a single pyroelectric detector design can be used for different wavelengths of radiation, simply by equipping the package with different windows coated with filters according to

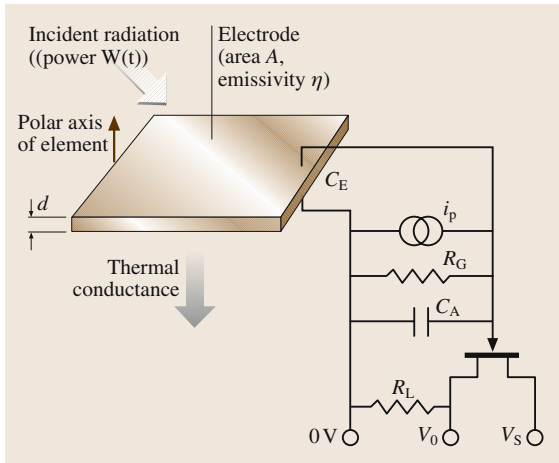


Fig. 27.17 Schematic diagram of a pyroelectric IR sensor

the radiation it is desired to sense. Such devices are widely used in spectroscopic sensors for, e.g. automotive exhaust-gas analysis. However, by far the widest application of pyroelectric detectors is in the sensing of IR in the wavelength range 8–14 μm . The reasons for this are that there is an atmospheric window of low absorption and that objects with temperatures in the range of room temperature (300 K) emit most of their radiative energy in that waveband. However, all of the semiconductor detectors of 8–14 μm IR need to be cooled to about 77 K to work efficiently. On the other hand, pyroelectric detectors work perfectly well uncooled. Cheap detectors are readily available for under \$1, and have thus found their way into a host of applications, notably detectors of people in such things as intruder alarms and remote light switches. Pyroelectric arrays have been developed in which a plane of pyroelectric material is interfaced to a silicon chip bearing a two-dimensional array of FET amplifiers and a set of switches which can be used to multiplex the pyroelectric signals onto a single output. These arrays have been used for uncooled thermal imagers with performances comparable to those that can be achieved with cooled semiconductor devices [27.32].

The requirements for the active materials in a pyroelectric device has been described by, for example, Whatmore [27.20], Whatmore and Watton [27.32]. There are two time constants that determine the frequency response of a pyroelectric device: the electrical time constant determined by $R_G(C_E + C_A)$ (see Fig. 27.17) and the thermal time constant determined by H/G where H is the heat capacity of the pyroelectric element and G the thermal conductance from the pyroelectric element

Table 27.7 Piezoelectric properties of PMN-PT and PZN-PT single-crystal materials [27.33] compared with those of a commercial soft-piezoelectric ceramic HD3203 [27.34]

Property	Composition	PZNT9307	PZNT9109	HD3203
$\epsilon_{33}^T/\epsilon_0$		3500	3000	3370
T_{RT} ($^\circ\text{C}$)		100	70	NA
T_C ($^\circ\text{C}$)		165	175	195
k_t (%)		58	56	0.536
k_{33} (%)		84	82	0.763
$N_{33,a}$		1230	1380	2000
d_{33} (pC/N)		1700	1900	564

to the environment. The capacitance of the element (C_E) is determined by the permittivity (ϵ) of the pyroelectric material while the heat capacity is determined by its specific heat per unit volume (c'). The pyroelectric current is proportional to the material's pyroelectric coefficient ($p = dP_s/dT$). A full treatment of the physics of device operation shows that detector performance (as determined by specific detectivity) is proportional to one of three basic figures-of-merit (FOMs), which combine the pyroelectric, dielectric and thermal properties of the pyroelectric material together. If the pyroelectric element capacitance is small in comparison with the amplifier capacitance (C_A), then the appropriate FOM is $F_I = p/c'$. If $C_E \gg C_A$ and the noise in the device is dominated by the amplifier noise sources, then the appropriate FOM is $F_V = p/c'\epsilon$. If the noise is dominated by alternating-current (AC) Johnson noise in the pyroelectric element then the appropriate FOM is $F_D = p/c'\sqrt{\epsilon \tan \delta}$, where $\tan \delta$ is the dielectric loss tangent of the pyroelectric material. Note that it is very important to measure the dielectric constant and loss at a frequency relevant to the device use. As most pyroelectric devices are used in the frequency range 0.1–100 Hz, and dielectric loss is usually much greater at < 100 Hz than at 1 kHz, a low-frequency measurement is essential. Table 27.7 lists the pyroelectric properties of several ferroelectric materials. Note that many of the literature papers on the subject only quote 1-kHz dielectric measurements. The differences between the FOM for the materials where high- and low-frequency data are available are clear. It can be seen that the TGS family gives the largest figures of merit. However, these materials are water soluble and have relatively low Curie temperatures and so they are only still used in the high-performance devices for instruments such as in Fourier-transform infra-red (FTIR) spectrometers. LiTaO₃ (LT) is a good pyroelectric material with relatively low permittivity.

It is inert and relatively easy to handle. It is used in many single-element detectors. Ceramic materials such as those based on modified PbZrO_3 (Mod PZ) or modified PbTiO_3 (Mod PT) are widely used in low-cost detectors. Note that, although their FOM are worse than LT, they are low cost and their performance is perfectly adequate for many applications. The **PVDF** family (represented here by a P(VDF/TrFE70/30 copolymer) have good F_V , but relatively low F_D values compared with the other materials listed. They tend to be used in large-area detectors because of their low permittivities, low cost and the fact that they are readily made in very thin films with low thermal mass, which is an advantage in some circumstances. They have been demonstrated in linear arrays. It is advantageous for very small-area detectors (say $< 100 \mu\text{m}$ square), such as those used in arrays, for the pyroelectric material to have a rel-

atively high permittivity (a few hundred) so that the detector can have a capacitance that matches the input capacitance of the **FET** amplifier (usually ca. 1 pF). The ceramic materials are well suited to this application for that reason. There has been considerable research into the use in thermal imaging arrays of ferroelectrics with T_C close to room temperature under an applied bias field, which will provide an induced pyroelectric effect well above the normal T_C . This has been called dielectric bolometer mode of operation and the best materials researched for this have been PST and $(\text{Ba}_x\text{Sr}_{1-x})\text{TiO}_3$ solid solutions with $x \approx 0.35$. These materials have relative permittivities of > 1000 under the operational conditions and very high pyroelectric coefficients which can give an effective F_D some three times greater than can be achieved using conventional pyroelectric ceramics [27.35].

References

- 27.1 G. Busch: *Ferroelectrics* **74**, 267 (1987)
- 27.2 A. F. Devonshire: *Phil. Mag.* **40**, 1040 (1949)
- 27.3 M. E. Lines, A. M. Glass: *Principles and Applications of Ferroelectric Materials* (Clarendon, Oxford 1977)
- 27.4 V. M. Goldschmidt: *Geochemistry* (Oxford Univ. Press, Oxford 1958)
- 27.5 R. D. Shannon, C. T. Prewitt: *Acta Cryst. B* **25**, 925 (1969)
- 27.6 W. J. Merz: *Phys. Rev.* **76**, 1221 (1949)
- 27.7 J. M. Herbert: *Ceramic Dielectrics and Capacitors* (Gordon Breach, Philadelphia 1985)
- 27.8 B. Jaffe, W. R. Cook Jr., H. Jaffe: *Piezoelectric Ceramics* (Academic, New York 1971)
- 27.9 R. W. Whatmore: Ph.D. Thesis, Cambridge University, UK (1977)
- 27.10 R. W. Whatmore, A. M. Glazer: *J. Phys. C: Solid State Phys.* **12**, 1505 (1979)
- 27.11 B. Noheda, J. A. Gonzalo, A. C. Caballero, C. Moure, D. E. Cox, G. Sirane: *Ferroelectrics* **237**, 541 (2000)
- 27.12 G. A. Smolenskii, V. A. Isupov, A. A. Agranovskaya, S. N. Popov: *Fiz. Tverd. Tela* **2**, 2906 (1960)
- 27.13 N.-H. Chan, D. M. Smyth: *J. Am. Ceram. Soc.* **67**, 285 (1984)
- 27.14 N.-H. Chan, R. K. Sharma, D. M. Smyth: *J. Am. Ceram. Soc.* **65**, 168 (1981)
- 27.15 S. C. Abrahams, E. Buehler, W. C. Hamilton, S. J. Laplace: *J. Phys. Chem. Solids* **34**, 521 (1973)
- 27.16 P. B. Jamieson, S. C. Abrahams, J. L. Bernstein: *J. Chem. Phys.* **48**, 5048 (1968)
- 27.17 M. E. Hagerman, K. R. Poeppelmeier: *Chem. Mater.* **7**, 602 (1995)
- 27.18 Y. J. Ding, X. D. Mu, X. H. Gu: *J. Non-Lin. Opt. Phys. Mater.* **9**, 21 (2000)
- 27.19 D. Madgy, S. F. Ahsan, D. Kest, I. Stein: *Arch. Otolaryng. Head Neck Surg.* **127**, 47 (2001)
- 27.20 R. W. Whatmore: *Rep. Prog. Phys.* **49**, 1335 (1986)
- 27.21 H. Kawai: *Jpn. J. Appl. Phys.* **8**, 967 (1969)
- 27.22 J. G. Bergman, J. H. McFee, G. R. Crane: *Appl. Phys. Lett.* **18**, 203 (1971)
- 27.23 J. A. Ghambaryan, R. Guo, R. K. Hovsepyan, A. R. Poghosyan, E. S. Vardanyan, V. G. Lazaryan: *J. Optoelectron. Adv. Mater.* **5**, 61 (2003)
- 27.24 A. Dabkowski, H. A. Dabkowska, J. E. Greedan, W. Ren, B. K. Mukherjee: *J. Cryst. Growth* **265**, 204–213 (2004)
- 27.25 R. Clarke, R. W. Whatmore: *J. Cryst. Growth* **33**, 29–38 (1976)
- 27.26 R. A. Dorey, R. W. Whatmore: *J. Electroceram.* **12**, 19 (2004)
- 27.27 N. Neuman, R. Köhler: *Proc. SPIE* **2021**, 35 (1993)
- 27.28 C. Pas de Araujo, J. F. Scott, G. W. Taylor: *Ferroelectric Thin Films: Synthesis and Basic Properties* (Gordon Breach, Princeton 1996)
- 27.29 J. F. Nye: *Physical Properties of Crystals, Their Representation by Tensors and Matrices*, (Oxford Univ. Press, Oxford 1957)
- 27.30 ANSI/IEEE Standard on Piezoelectricity: *IEEE Trans. UFFC* **43**, 717 (1996)
- 27.31 MIL-STD-1376B(SH) (1995) US Navy Military Standard for Piezoelectric Ceramics
- 27.32 R. W. Whatmore, R. Watton: *Pyroelectric Materials and Devices*. In: *Infrared Detectors and Emitters: Materials and Devices*, ed. by P. Capper, C. T. Elliott (Chapman Hall, New York 2000) pp. 99–148
- 27.33 Y. Hosono, K. Harada, T. Kobayashi, K. Itsumi, M. Izumi, Y. Yamashita, N. Ichinose: *Jpn. J. Appl. Phys.* **41**, 7084–8 (2002)

- 27.34 S. Sherrit, H. D. Wiederick, B. K. Mukherjee: Proc. SPIE **3037**, 158 (1997)
- 27.35 R. W. Whatmore, P. C. Osbond, N. M. Shorrocks: Ferroelectrics **76**, 351 (1987)

28. Dielectric Materials for Microelectronics

Dielectrics are an important class of thin-film electronic materials for microelectronics. Applications include a wide swathe of device applications, including active devices such as transistors and their electrical isolation, as well as passive devices, such as capacitors. In a world dominated by Si-based device technologies, the properties of thin-film dielectric materials span several areas. Most recently, these include high-permittivity applications, such as transistor gate and capacitor dielectrics, as well as low-permittivity materials, such as inter-level metal dielectrics, operating at switching frequencies in the gigahertz regime for the most demanding applications.

This chapter provides a survey of the various dielectric material systems employed to address the very substantial challenge associated with the scaling Si-based integrated circuit technology. A synopsis of the challenge of device scaling is followed by an examination of the dielectric materials employed for transistors, device isolation, memory and interconnect technologies. This is presented in view of the industry roadmap which captures the consensus for device scaling (and the underlying economics) – the International Technology Roadmap for Semiconductors. Portions

28.0.1	The Scaling of Integrated Circuits.	625
28.0.2	Role of Dielectrics for ICs	629
28.1	Gate Dielectrics	630
28.1.1	Transistor Structure	630
28.1.2	Transistor Dielectric Requirements in View of Scaling .	630
28.1.3	Silicon Dioxide	635
28.1.4	Silicon Oxynitride: SiO _x N _y	641
28.1.5	High- κ Dielectrics	643
28.2	Isolation Dielectrics	647
28.3	Capacitor Dielectrics	647
28.3.1	Types of IC Memory	647
28.3.2	Capacitor Dielectric Requirements in View of Scaling	648
28.3.3	Dielectrics for Volatile Memory Capacitors	648
28.3.4	Dielectrics for Nonvolatile Memory	649
28.4	Interconnect Dielectrics	651
28.4.1	Tetraethoxysilane (TEOS)	651
28.4.2	Low- κ Dielectrics	651
28.5	Summary	653
	References	653

of the survey presented here are selected from work previously published by the author [28.1–3].

This chapter considers the role of dielectric materials in microelectronic devices and circuits and provides a survey of the various materials employed in their fabrication. We will examine the impact of scaling on these materials, and the various materials utilized for their dielectric behavior. Extensive reviews are available on the device characteristics for the reader to consult [28.4–7]. We will primarily confine the discussion here to Si-based microelectronic circuits.

Dielectric materials are an integral element of all microelectronic circuits. In addition to their primary function of electrical isolation of circuit and device components, these materials also provide useful chemical and interfacial properties. The material (and result-

ing electrical) properties of dielectrics must also be considered in the context of the thin films used in semiconductor microelectronics, as compared to bulk properties. The dimensions of these dielectric thin films are determined by the device design of the associated integrated circuit technology, and these dimensions decrease due to a calculated design process called *scaling*.

28.0.1 The Scaling of Integrated Circuits

The ability to reduce the size of the components of integrated circuits (ICs), and therefore the circuits themselves, has resulted in substantial improvements in device and circuit speeds over the last 30 years. Equally

important, this calculated size reduction permits the fabrication of a higher density of circuits per unit area on semiconductor substrates. The economic implication of this scaling was captured by *G. Moore* more than 40 years ago [28.8].

Moore's Law

Moore observed that the minimum cost of manufacturing integrated circuits per component actually decreases with increasing number of IC components, and thus with greater circuit functionality and computing power. This is obviously an important economic driving force, as the ability to increase the number of circuits per unit area would lead to a lower minimum cost, and thus higher market demand and more potential profit. Moreover, Moore noted that the rate of increase in the number of components for a given circuit function roughly doubled each year in the early 1960s, and predicted that it would continue to do so through 1975. In 1975, *Moore* revised this estimate of doubling time to 24 months due to the anticipated complexity of circuits [28.9].

The semiconductor industry has generally confirmed (and aligned to meet) these extrapolations, often referred to as Moore's law, over the last 30 years. The extrapolation is often analyzed in the semiconductor industry, and the doubling period, which has varied between 17 and 32 months over the life of the industry, is now roughly 23 months [28.10]. Indeed the cost per transistor has decreased from ≈ 5 \$/transistor in 1965 to less than 10^{-6} \$/transistor today [28.11]. Current advanced Si IC production technology results in the fabrication of well over 500 000 000 transistors on a microprocessor chip.

Technology Roadmap

The contemporary industry analysis encompassing this observation is presented in the International Technology Roadmap for Semiconductors (ITRS) where the extrapolations of the future technological (and economic) requirements for the industry are annually updated [28.12]. The current scaling trends indicate that the compound annual reduction rate (CARR) in device dimensions is currently consistent with the following equation:

$$\text{CARR}(T) = \frac{1}{2} \left(\frac{1}{2^T} \right) - 1, \quad (28.1)$$

where T is the technology node cycle time measured in years. Thus in two years, the rate of reduction is -15.9% . This corresponds to a scaling factor of $\approx 0.7 \times$ from a given technology node to the next, or a factor of

$\approx 0.5 \times$ over the time of two technology nodes. Recent reviews and predictions for the limits of scaled integrated circuit technology continue to be available [28.13, 14].

Table 28.1 provides selected scaling targets from the ITRS through 2010 [28.12]. As may be seen, the industry roadmap now segregates scaling targets among three categories: microprocessor (MPU)/application-specific IC (ASIC) applications, low-operational-power applications, and low-stand-by-power applications. Higher-performance technologies, such as MPU/ASIC applications, require aggressive scaling, while low-stand-by-power applications require less-aggressive scaling. A key criteria to enable these technologies is the associated power management in the *on* and *off* state, hence leakage current remains an important distinction amongst the various roadmap applications.

Performance and Scaling

The concomitant reduction in device dimensions, such as transistor channel length, associated with increasing the number of components per unit area has resulted in a significant increase in processing performance – the speed at which computations can be done. For example, as scaling reduces the distance that carriers must travel in a transistor channel, the response time of the transistor as a digital switch also decreases (as long as sufficient mobility is maintained). Scaling has resulted in the speeds of microprocessors increasing from 25–50 MHz clock frequencies in the early 1990s ($\approx 10^6$ transistors) to 2.2 GHz in 2003 ($\approx 10^8$ transistors) [28.15]. Current predictions suggest that 20-GHz frequencies are possible for complementary metal oxide semiconductor (CMOS)-based microprocessors incorporating 10^9 transistors [28.11].

The various dielectric materials associated with the components impacts the overall performance of the corresponding IC technology. In the case of transistors, the gate dielectric is integral to the performance of transistor electrical characteristics such as the drive current I_d . The interconnection performance of circuit elements is influenced by the dielectric material that isolates the various metal interconnection lines through the line-to-line capacitance. Memory elements incorporate very-low-leakage dielectric materials for charge-storage purposes. The behavior of these materials with component size reduction is thus an important design consideration in IC technology.

Figure 28.1 shows the dependence of the delay time as a function of technology node (scaling) associated with conventional CMOS transistor gates interconnected with metal lines that are isolated with different di-

Table 28.1 Selected scaling targets from the ITRS roadmap [28.12]

Year of production	2003	2004	2005	2006	2007	2008	2009	2010
Technology node	100 nm	90 nm	80 nm	70 nm	65 nm	57 nm	50 nm	45 nm
MPU/ASIC applications								
Physical gate length, MPU/ASIC (nm)	45	37	32	28	25	22	20	18
Equivalent oxide thickness, T_{ox} (nm)	1.3	1.2	1.1	1.0	0.9	0.8	0.8	0.7
Drive voltage V_{dd} (V)	1.2	1.2	1.1	1.1	1.1	1	1	1
Gate dielectric leakage ($\mu A/\mu m$)	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.3
Gate dielectric leakage (A/cm^2)	222	450	521	595	933	1061	1167	1852
Low operating power								
Physical gate length, low operating power (nm)	65	53	45	37	32	28	25	22
T_{ox} (nm)	1.6	1.5	1.4	1.3	1.2	1.1	1.0	0.9
V_{dd} (V)	1.00	0.90	0.90	0.90	0.80	0.80	0.80	0.70
Gate dielectric leakage (nA/ μm)	0.33	1.00	1.00	1.00	1.67	1.67	1.67	2.33
Gate dielectric leakage (A/cm^2)	0.51	1.89	2.22	2.70	5.21	5.95	6.67	10.61
Low stand-by power								
Physical gate length, low stand-by power (nm)	75	65	53	45	37	32	28	25
T_{ox} (nm)	2.2	2.1	2.1	1.9	1.6	1.5	1.4	1.3
V_{dd} (V)	1.2	1.2	1.2	1.2	1.1	1.1	1.1	1.0
Gate dielectric leakage (pA/ μm)	3	3	5	7	8	10	13	20
Gate dielectric leakage (A/cm^2)	0.004	0.005	0.009	0.015	0.023	0.031	0.048	0.080

electronics [28.16]. The total delay time associated with the circuit response for a given technology node is a combination of the temporal response of the transistor gate (influenced by the gate capacitance) as well as the propa-

gation through the interconnecting wire lines associated with the circuit [influenced by the wire resistance and the line-to-line capacitance, i.e. a resistive-capacitive (RC) delay] [28.16, 17]. Changing both the interconnect wire material (from Al to Cu) to decrease the line resistivity as well as decreasing the dielectric constant in the material isolating these lines (from SiO_2 to a lower dielectric constant material) to decrease line-to-line parasitic capacitance minimizes the overall RC delay time response. Of course, the challenge in adopting this solution was the successful integration of these material changes into a high-volume fabrication line, which took the investment of considerable time and effort.

In order to compare the performance of different CMOS technologies, one can consider a performance metric that captures the dynamic response (i.e. charging and discharging) of the transistors associated with a specific circuit element to the supply voltage provided to the element at a representative (clock) frequency [28.1, 18, 19]. A common element employed to examine such switching-time effects in a new transistor design technology node is a CMOS inverter. This circuit element is shown in Fig. 28.2 where the input signal is attached to the gates and the output signal is connected to both the n-type MOS (NMOS) and p-type MOS (PMOS) transistors associated with the CMOS stage. The switching time is limited by both the fall time required to discharge

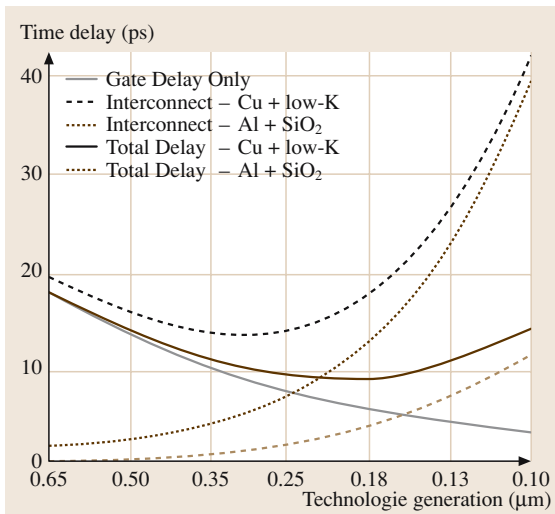


Fig. 28.1 Time delay contributions for integrated circuits. The resultant delay is dependent upon the circuit architecture as well as the material constituents. After [28.16] (© 2004 IEEE, with permission)

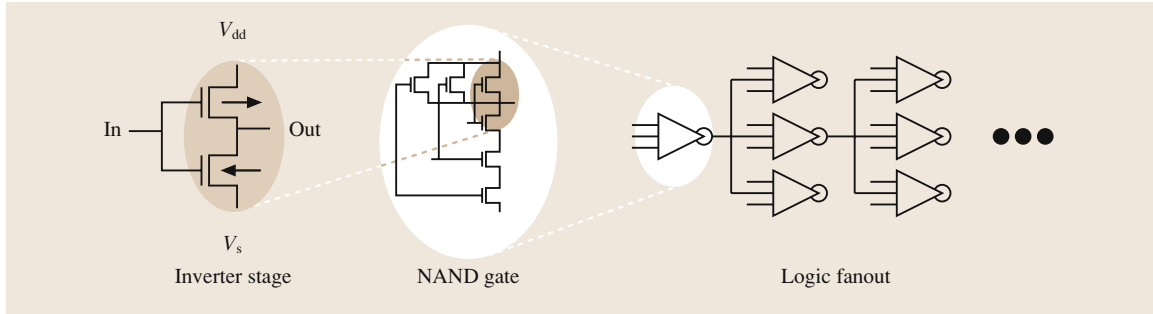


Fig. 28.2 Example circuits employed to evaluate the performance of a transistor technology. For example, elements are combined to simulate processor performance adequately. After [28.1] (AIP, with permission© 2001)

the load capacitance by the n-type field-effect transistor (n-FET) drive current and the rise time required to charge the load capacitance by the p-FET drive current. That is, the switching response times τ are given by:

$$\tau = \frac{C_{\text{LOAD}} V_{\text{dd}}}{I_d}, \quad \text{where}$$

$$C_{\text{LOAD}} = FC_{\text{GATE}} + C_j + C_i, \quad (28.2)$$

and C_j and C_i are the parasitic junction and local inter-connection capacitances, respectively. Of course, as one considers the combination of transistors to produce various circuit logic elements to enable computations (viz. a *processor*), it is also clear that the manner in which the transistors are interconnected also plays an important role on the overall performance of the circuit. The *fan out* for interconnected devices is given by the factor F . Ignoring delay in the gate electrode response, as $\tau_{\text{GATE}} \ll \tau_{n,p}$, the average switching time is therefore:

$$\bar{\tau} = \frac{\tau_p + \tau_n}{2} = C_{\text{LOAD}} V_{\text{dd}} \left(\frac{1}{I_d^n + I_d^p} \right). \quad (28.3)$$

The load capacitance in the case of a single CMOS inverter is simply the gate capacitance if one ignores parasitic contributions such as junction and interconnect capacitance. Hence, an increase in I_d is desirable to reduce switching speeds. For more realistic estimates of microprocessor performance, the load capacitance is connected (*fanned out*) to other inverter elements in a predetermined fashion. When coupled with other NMOS/PMOS transistor pairs in the configuration shown in Fig. 28.2, one can create a logic NAND gate, which can be used to investigate the dynamic response of the transistors and thus examine their performance under such configurations. For example, a fan out $F = 3$ can be employed in microprocessor performance estimates, as shown in Fig. 28.2.

One can then characterize the performance of a circuit technology (based on a particular transistor structure) through this switching time. To do this, various *figures of merit* (FOM) have been proposed that incorporate parasitic capacitance as well as the influence of gate sheet resistance on the switching time [28.18–20]. For example, a common FOM employed is related to (28.3) simply by

$$\text{FOM} \cong \frac{1}{\bar{\tau}} = \frac{2}{\tau_p + \tau_n}. \quad (28.4)$$

In the case where parasitic capacitance and resistance effects are ignored, it is easily seen then that an increase in the device drive current I_d results in a decrease in the switching time and thus an increase in the FOM value (performance). However, the incorporation of parasitic effects results in the limitation of FOM improvement, despite an increase in the gate dielectric capacitance. All of these issues depend critically upon the materials constituents of the integrated circuit, and dielectrics are an important component [28.1].

Impact of Scaling on Dielectrics

The scaling associated with the device dimensions that constitute an integrated circuit impacts on virtually all film dimensions such as thickness, including the dielectrics employed in the circuit. Various scaling methodologies have been applied since the pioneering work by Dennard and coworkers [28.21, 22]. These various methods have been recently summarized [28.5, 23]. For example, following the work by Dennard et al., the scaling of the transistor from one node N to the next $N + 1$ under a constant-electric-field condition would result in a relationship between the *new* transistor gate (or channel) length $L_{N+1} = L_N/\alpha$, where $\alpha > 1$ is the constant scaling factor (Fig. 28.3). Other dimensions, such as the gate dielectric thickness and the gate width,

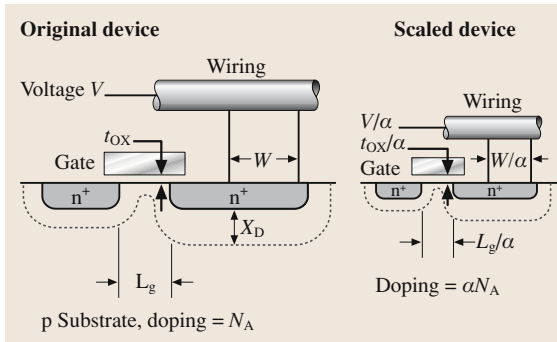


Fig. 28.3 Scaling methodology for integrated circuit technology. After [28.23] (© 2004 IEEE, with permission)

scale similarly. The voltage would also scale according to $V_{N+1} = V_N/\alpha$ in this scenario as well. Table 28.1 provides the drive voltage scaling (V_{dd}) anticipated by the ITRS roadmap, depending upon the application.

This scaling approach has several shortcomings, including the reduction of power-supply voltages with each node, raising problems with circuit power-supply compatibility. So, for earlier technology nodes ($L \geq 0.35 \mu\text{m}$), the power supply was kept constant while the device dimensions were scaled as described above. Eventually, the resultant higher electric fields associated with this scaling approach led to reliability

concerns for the gate dielectric, mobility degradation and hot-carrier effects. For further scaling, the power supply was scaled at a different factor (such as $\alpha^{1/2}$) to compensate for such effects. For modern devices with *deep-submicron* gate length, other techniques are used to scale from node to node. Nevertheless, for CMOS technology over the last ≈ 25 years, it has been noted that the gate dielectric thickness can be related to the channel length as shown in Fig. 28.4 for a wide variety of device technologies [28.24].

The critical point in the context of dielectric films is that the dimensions of the devices, and therefore the dimensions (thicknesses) of the dielectric layers, are significantly reduced as scaling proceeds (Table 28.1). At some point, the materials properties associated with the dielectric in a component will be predicted to no longer provide the desirable electrical behavior attained in previous nodes, thus stimulating research on new dielectric materials. In the case of gate dielectrics, scaling results in thinner layers to a point where quantum-mechanical tunneling becomes an issue. For memory capacitor dielectrics, a specific capacitance density is required, regardless of capacitor dimensions, for a reliable memory element. Indeed, scaling impacts on the selection and integration of numerous constituents of electronic devices including gate electrode materials, source/drain junction regions, contacts, interconnect dielectrics.

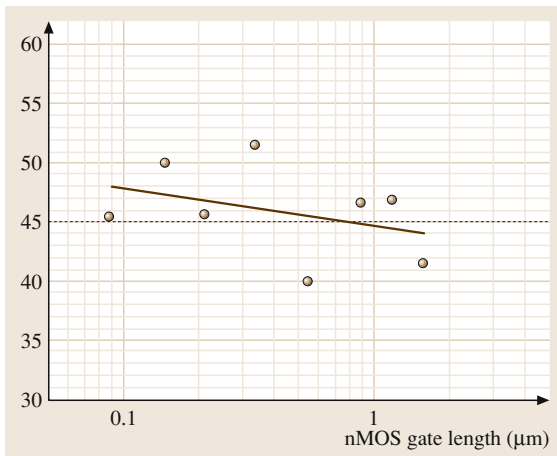


Fig. 28.4 Scaling trends of the transistor gate length and the gate insulator thickness for various generations of integrated technologies. The *solid line* is a linear fit to the data. The *dashed line* represents the approximate scaling trend between the gate dielectric thickness and gate length observed over several technology generations: $L_G = 45 \times t_{ox}$. After [28.24], with permission

28.0.2 Role of Dielectrics for ICs

Dielectrics are pervasive throughout the structure of an integrated circuit. Applications include gate dielectrics, tunneling oxides in memory devices, capacitor dielectrics, interconnect dielectrics and isolation dielectrics, as well as sacrificial or masking applications during the circuit fabrication process.

The basic properties utilized in dielectric materials include their structure and the resultant polarizability behavior. As noted in Fig. 28.5, there are several mechanisms to consider in the polarization of materials in the presence of an electric field. For nonmetallic (insulating) solids in CMOS applications, there are two main contributions of interest to the dielectric constant which give rise to the polarizability: electronic and ionic dipoles. Figure 28.5 illustrates the frequency ranges where each contribution to the real part of the complex relative dielectric permittivity ϵ_r dominates, as well as the imaginary part of the complex permittivity (dielectric losses), ϵ_r'' [28.25]. The region of interest to CMOS applications includes the shaded area where switching frequencies are in the GHz range at the present time.

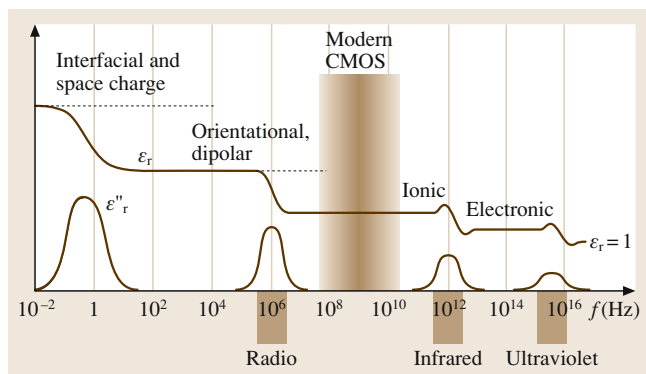


Fig. 28.5 Schematic of the dependence of dielectric permittivity on frequency. After [28.25], with permission

In general, atoms with a large ionic radius (i.e. high atomic number, Z) exhibit more electron dipole response to an external electric field, because there are more electrons to respond to the field (electron screening effects also play a role in this response). This electronic con-

tribution tends to increase the permittivity of materials with higher- Z atoms.

The ionic contribution to the permittivity can be much larger than the electronic portion in cases such as perovskite crystals of $(\text{Ba},\text{Sr})\text{TiO}_3$ (BST) and $(\text{Pb},\text{Zr})\text{TiO}_3$ (PZT), which exhibit ferroelectric behavior below the Curie point. In these cases, Ti ions in unit cells throughout the crystal are uniformly displaced in response to an applied electric field (for the case of ferroelectric materials, the Ti ions reside in one of two stable, non-isosymmetric positions about the center of the Ti–O octahedra). This displacement of Ti ions causes an enormous polarization in the material, and thus can give rise to very large dielectric constants in bulk films of 2000–3000, and has therefore been considered for dynamic random-access memory (DRAM) capacitor applications [28.26, 27]. Since ions respond more slowly than electrons to an applied field, the ionic contribution begins to decrease at very high frequencies, in the infrared range of $\approx 10^{12}$ Hz, as shown in Fig. 28.5.

28.1 Gate Dielectrics

28.1.1 Transistor Structure

A cross-sectional schematic of the structure of a metal–oxide–semiconductor (MOS) field-effect transistor (FET) and a modern CMOS transistor, consisting of the n-FET and p-FET pair, is shown in Fig. 28.6. The source- and drain-region dopant profiles reflect modern planar CMOS technologies where an extended, doped region is produced under the gate region in the channel. Additionally, so-called *halo* or *pocket* dopant implantation regions are also shown. These dopant profile approaches have been incorporated in recent years in an effort to permit transistor channel scaling and yet maintain performance [28.28].

As can be seen, there exist several regions of the device which require dielectric materials to enable useful transistor operation. Both n-MOS and p-MOS transistors are isolated with a dielectric, typically deposited SiO_2 in most modern device technologies, which is deposited in trench structures. This isolation technique is called *shallow trench isolation* (STI) as the associated trench depth is $\leq 0.5\ \mu\text{m}$. Earlier generations of CMOS (and often devices under research) also utilized the so-called local oxidation of silicon (LOCOS) isolation approach [28.29, 30]. So-called spacer dielectrics (typically SiO_2 , SiO_xN_y , or SiN_x) are also used around

the transistor *gate stack* for isolation and implantation-profile control. The gate stack is defined here as the films and interfaces comprising the gate electrode, the underlying gate dielectric, and the channel region. The gate dielectric, typically SiO_2 or SiO_xN_y for transistors currently in production, electrically isolates the gate electrode from the underlying Si channel region while allowing the modulation of the carrier flow in the channel. As we shall see, the interface between the gate dielectric and the channel regions is particularly important in regard to device performance.

28.1.2 Transistor Dielectric Requirements in View of Scaling

In addition to the thickness reduction associated with scaling, the gate dielectric must adhere to several requirements that are dependent upon the specific product application, including adequate drive current, suitable capacitance, minimal leakage current, and reliable performance. Table 28.1 provides some selected values describing the ITRS 2003 roadmap.

Drive Current

The improved performance associated with the scaling of logic device dimensions can be seen by consider-

ing a simple model for the drive current associated with a FET [28.1, 3, 18, 20]. In the gradual channel approximation, the drive current can be written as:

$$I_d = \frac{W}{L} \mu C_{\text{inv}} \left(V_G - V_T - \frac{V_D}{2} \right) V_D, \quad (28.5)$$

where W is the width of the transistor channel, L is the channel length, μ is the channel carrier mobility (assumed to be constant in this analysis), C_{inv} is the capacitance density associated with the gate dielectric when the underlying channel is in the inverted state, V_G and V_D are the voltages applied to the transistor gate and drain, respectively, and the threshold voltage is given by V_T . It can be seen that in this approximation the drain current is proportional to the average charge across the channel (with a potential $V_D/2$) and the average electric field (V_D/L) along the channel direction. Initially, I_D increases linearly with V_D and then eventually saturates to a maximum when $V_{D,\text{sat}} = V_G - V_T$ to yield:

$$I_{D,\text{sat}} = \frac{W}{L} \mu C_{\text{inv}} \frac{(V_G - V_T)^2}{2}. \quad (28.6)$$

A goal here is to increase the saturation drive current as much as possible for a given transistor technology design. The term $(V_G - V_T)$ is limited in range due to reliability concerns as too large a V_G would create an undesirable, high electric field across the gate dielectric. Moreover, practical operation of the device at room temperature (or above) requires that V_T cannot easily be reduced below about $8 \times k_B T \cong 200$ meV ($k_B T \approx 25$ meV at room temperature). Elevated operating temperatures could therefore cause statistical fluctuations in thermal energy, which would adversely affect the desired V_T value. Thus, even in this simplified approximation, one is left to pursue a reduction in the channel length, increasing the width, increasing the gate dielectric capacitance, or a carefully engineered combination of all of these in order to increase $I_{D,\text{sat}}$. Note that none of these dimensions are, in reality, independent and so substantial resources are utilized to simulate device performance in order to maximize the performance prior to the fabrication of a representative transistor [28.5, 30, 32].

Gate Capacitance

In the case of increasing the inversion capacitance to improve I_D , consider the gate stack structure idealized as a parallel-plate capacitor (ignoring quantum-mechanical and depletion effects from a Si substrate and

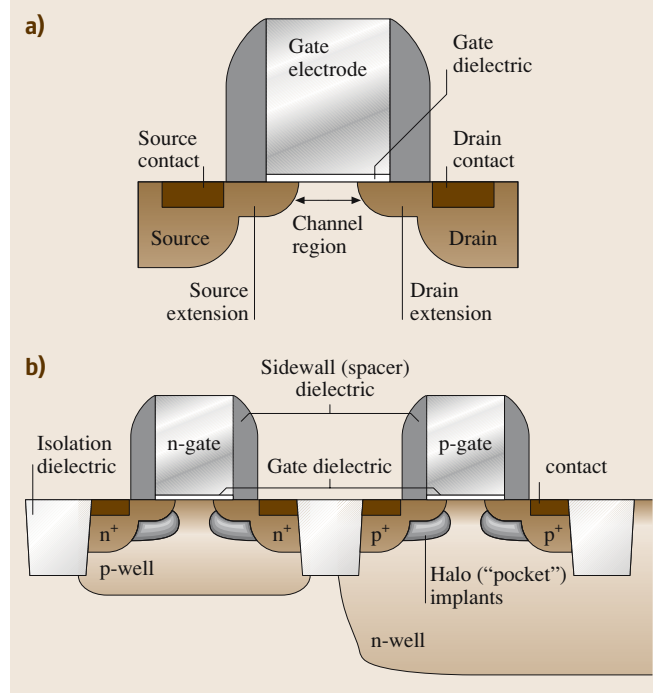


Fig. 28.6a,b Important regions of (a) a MIS field-effect transistor, and (b) a planar CMOS transistor structure. After [28.31] (© 2004 IEEE, with permission)

gate [28.33]),

$$C = \frac{\kappa \epsilon_0 A}{t}, \quad (28.7)$$

where κ is the dielectric constant of the material, ϵ_0 is the permittivity of free space ($= 8.85 \times 10^{-3}$ fF/ μm), A is the area of the capacitor, and t is the thickness of the dielectric between the capacitor electrode plates. (Note that the relative permittivity of a material is often given by ϵ or ϵ_r , such as with the expression $C = \epsilon \epsilon_0 A/t$. Note that the relation between κ and ϵ varies depending on the choice of units (e.g. when $\epsilon_0 = 1$), but since it is always the case that $\kappa \approx \epsilon$, we shall assume here that $\kappa = \epsilon$.) It is clear that transistor designs utilizing planar CMOS scaling approaches discourage increasing the area of the capacitor, and so one is left to considering the reduction of the dielectric thickness in order to increase the capacitance.

In order to compare alternative dielectric materials which exhibit a dielectric constant higher than the standard for the industry, SiO_2 , this expression for C can also be rewritten in terms of t_{eq} (i.e. equivalent oxide thickness) and $\epsilon \kappa_{\text{ox}}$ of the capacitor ($\kappa_{\text{ox}} = 3.9$

for the low-frequency dielectric constant of SiO_2). The term t_{eq} then represents the theoretical thickness of SiO_2 that would be required to achieve the same capacitance density as the dielectric (ignoring important issues such as leakage current and reliability). For example, if the capacitor dielectric is SiO_2 , $t_{\text{eq}} = 3.9 \varepsilon_0(A/C)$, and so a capacitance density of $C/A = 34.5 \text{ fF}/\mu\text{m}^2$ corresponds to $t_{\text{eq}} = 10 \text{ \AA}$.

Thus, the physical thickness of an alternative high- κ dielectric employed to achieve the equivalent capacitance density of $t_{\text{eq}} = 10 \text{ \AA}$ can be obtained from the expression

$$\frac{t_{\text{eq}}}{\kappa_{\text{ox}}} = \frac{t_{\text{high-}\kappa}}{\kappa_{\text{high-}\kappa}} \quad \text{or simply,}$$

$$t_{\text{high-}\kappa} = \frac{\kappa_{\text{high-}\kappa}}{\kappa_{\text{ox}}} t_{\text{eq}} = \frac{\kappa_{\text{high-}\kappa}}{3.9} t_{\text{eq}}. \quad (28.8)$$

From this expression, a dielectric with a relative permittivity of 16 therefore results in a physical thickness of $\approx 40 \text{ \AA}$, to obtain $t_{\text{eq}} = 10 \text{ \AA}$. The increase in the physical thickness of the dielectric impacts on properties such as the tunneling (leakage) current through the dielectric, and is discussed further below.

The industry scaling process clearly presents a major challenge for the core transistor gate dielectric as predictions call for a much thinner effective thickness for future alternative gate dielectrics: $t_{\text{eq}} \leq 1 \text{ nm}$ [28.12]. The interfacial regions between the gate electrode, dielectric and channel [in totality termed the metal–insulator–semiconductor (MIS) *gate stack* as shown schematically in Fig. 28.7] require careful attention, as they are particularly important in regard to transistor performance. These regions, less than approximately 0.5 nm thick, serve as a transition between the atoms associated with the materials in the gate electrode, gate dielectric and Si

channel, and can alter the overall capacitance of the gate stack, particularly if they have a thickness that is substantial relative to the gate dielectric. Additionally, these interfacial regions can be exploited to obtain desirable properties. The upper interface, for example, can be engineered in order to block boron out-diffusion from the poly-Si gate. The lower interface, which is in direct contact with the CMOS channel region, must be engineered to provide low interface trap densities (e.g. dangling bonds) and minimize carrier scattering (maximize channel carrier mobility) in order to obtain reliable, high-performance device. Mobility degradation, relative to that obtained using SiO_2 (or SiON) gate dielectrics, associated with the incorporation of high- κ gate dielectrics is an important issue currently under investigation and discussed further below.

Reactions at either of these interfaces during the device fabrication process can result in the formation of a significant interfacial layer that will likely reduce the desired gate stack capacitance. Additionally, any suitable interfacial layer near the channel must result in a low density of electrically active defects ($\lesssim 10^{11}/\text{cm}^2\text{eV}$ is often obtained for SiO_2) and avoid degradation of carrier mobility in the region near the surface channel.

The reduced capacitance can be seen by noting that the dielectric film that has undergone interfacial reactions results in a structure that essentially consists of several dielectric layers in series. If we suppose that the dielectric consists of two layers, then from electrostatics the total capacitance of two dielectrics in series is given by $1/C_{\text{tot}} = 1/C_1 + 1/C_2$, where C_1 and C_2 are the capacitances of the two layers. Thus, the lowest-capacitance layer will dominate the overall capacitance and also set a limit on the minimum achievable t_{eq} value. If we consider a dielectric stack structure such that the bottom layer (layer 1) of the stack is SiO_2 , and the top layer (layer 2) is the high- κ alternative gate dielectric, (28.8) is expanded (assuming equal areas) to:

$$t_{\text{eq}} = t_{\text{SiO}_2} + \frac{\kappa_{\text{SiO}_2}}{\kappa_{\text{high-}\kappa}} t_{\text{high-}\kappa}. \quad (28.9)$$

From (28.9), it is clear that the layer with the lower dielectric constant (SiO_2 in this case) limits the ultimate capacitance of the MIS gate stack.

Of course, the actual capacitance of a CMOS gate stack for ultra-large-scale integration (ULSI) devices does not scale simply with $1/t$ due to parasitics, quantum-mechanical confinement of carriers, and depletion effects. Indeed, consideration of these important effects can result in much confusion on the defini-

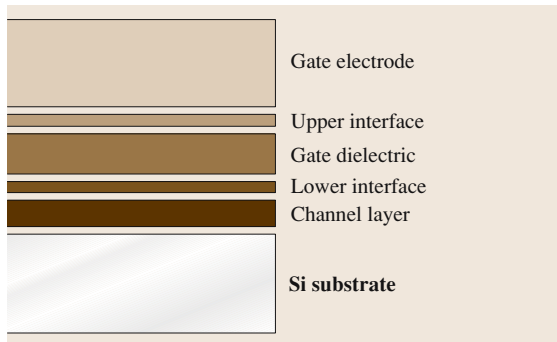


Fig. 28.7 Schematic of the important regions of the transistor gate stack. After [28.1] (AIP, with permission © 2001)

tion of dielectric thickness as extracted from electrical measurements.

Parasitic resistances and capacitances associated with various portions of the transistor structure can result in an overall degradation of performance as defined by a delay time figure of merit [28.1, 19]. There exist several materials issues associated with the control of such parasitics including source/drain dopant-profile control, gate/contact sheet resistance minimization, etc. which are beyond the scope of this chapter.

Quantum-mechanical confinement effects on carriers in the channel region occur as a result of the large electrical fields in the vicinity of the Si substrate surface. These fields quantize the available energy levels resulting in the displacement of the charge centroid from the interface into the Si and at energies above the Si conduction-band edge. The extent of the penetration of the charge centroid is dependent upon the biasing conditions employed for the metal–insulator–semiconductor (MIS) structure and, for accurate estimates of transistor drain-current performance, the inversion capacitance measurement provides accurate determination of the equivalent oxide thickness [28.33, 36, 37].

Current MIS gate-stack structures employ heavily doped polycrystalline Si (poly-Si) as the *metal* gate electrode for CMOS transistors. Poly-Si gates were introduced as a replacement for Al metal gates in the 1970s for CMOS integrated circuits and have the very desirable property of a tunable work function for both n-MOS and p-MOS transistors through the implantation of the appropriate dopant into the poly-Si gate electrode.

However, in deep-submicron scaled devices, the *poly-Si depletion effect* impacts the overall gate-stack capacitance significantly, as the scaling results in an increased sensitivity to the effective *electrical* thickness resulting from all of the gate-stack component films, including the gate electrode. The depletion effect (which occurs in any doped semiconductor) is a result of the decrease in the density of majority carriers near the poly-Si/dielectric interface upon biasing, resulting in an increased depletion width in the poly-Si [28.28, 38]. This results in a voltage drop across the depletion region, rendering a smaller voltage drop across the remaining portions of the gate stack (i.e. the dielectric and the substrate). As a result, the (low-frequency) capacitance under strong inversion biasing conditions is significantly lower than that obtained under accumulation (Fig. 28.8). Moreover, the inversion-layer charge in the substrate is reduced from that possible in the ideal case where no depletion layer exists in the gate (i.e. an ideal metal conductor for the gate electrode). This reduction in the

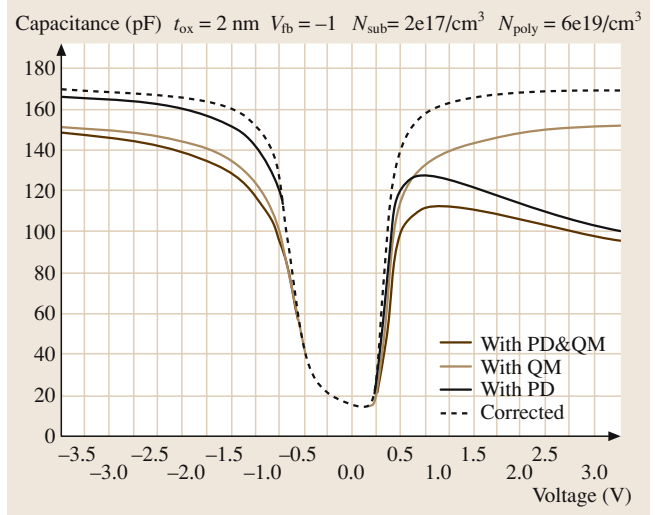


Fig. 28.8 Capacitance–voltage curves demonstrating the effects of poly-Si depletion and quantum-mechanical behavior for scaled transistors. N_{sub} and N_{poly} denote the dopant density for the substrate and poly-Si gate, respectively. After [28.34]

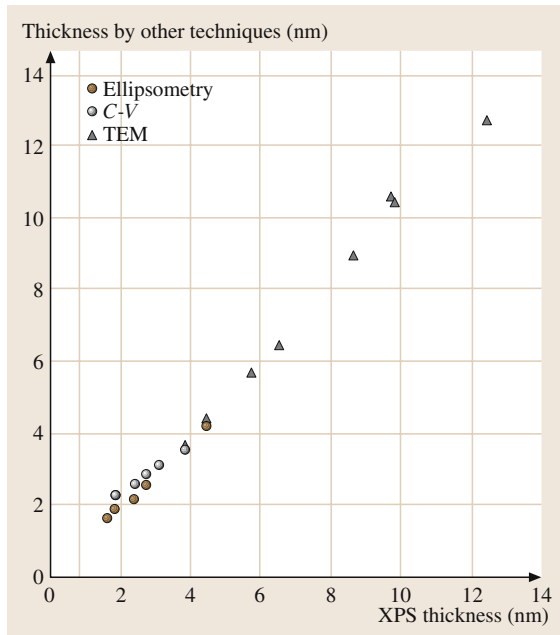


Fig. 28.9 Metrology results for various measurement techniques of the thickness of thin SiO₂ films. X-ray photoelectron spectroscopy (XPS) measured thicknesses are compared to thicknesses derived from high-resolution transmission electron microscopy (HRTEM), C–V and ellipsometry. After [28.35] (AIP, with permission © 1997)

inversion-layer charge in turn reduces the ideal transistor drive current. As the dielectric thickness is reduced in the scaling process, this effect becomes a serious limitation for transistor drive current.

To solve this problem, a return to metal gate electrode materials is now under development to eliminate any depletion region in the gate, and thus the interaction of any dielectric with various metal gate candidates is of interest. Compatibility of dielectric materials with metal gate materials is briefly discussed later in this chapter.

When a capacitance–voltage curve is properly corrected for these quantum-mechanical and dopant depletion effects, as seen in Fig. 28.8, one can accurately extract the equivalent oxide thickness (EOT), t_{eq} [28.34]. This thickness definition is in contrast to that derived directly from the raw data in a capacitance–voltage (C–V) measurement, which is termed the *capacitance equivalent thickness* (CET > EOT), or the *physical thickness*, which can be determined by non-electrical characterization methods such as ellipsometry or high-resolution transmission electron microscopy, as seen in Fig. 28.9 [28.34, 35]. Please note that thicknesses derived from quantum-mechanical-corrected accumulation capacitance data are also often abbreviated as t_{qm} .

Leakage Current

Of course, physically thinning the dielectric layer to the nanometer regime raises the prospect of quantum-mechanical tunneling through the dielectric – often referred to as *leakage*. To minimize gate leakage (tunneling) currents, quantum mechanics indicates that the gate dielectric must be sufficiently thick and have a sufficient energy barrier (band offset) to minimize the resultant tunneling current. Consider the band diagram shown in Fig. 28.10, where the electron affinity (χ) and gate

electrode work function (Φ_M) are defined (q is the charge). For electrons traveling from the Si substrate to the gate, this barrier is the conduction-band offset, $\Delta E_C \cong q[\chi - (\Phi_M - \Phi_B)]$; for electrons traveling from the gate to the Si substrate, this barrier is Φ_B . The effect of the height of the energy barrier and the film thickness on tunneling current can be seen by considering the expression for the direct tunneling current where electron transport occurs through a trapezoidal energy barrier [28.18]:

$$J_{dt} = \frac{A}{t_{diel}^2} \exp(-Bt_{diel}) \times \left\{ \left(\Phi_B - \frac{V_{diel}}{2} \right) \exp \left[\sqrt{\left(\Phi_B - \frac{V_{diel}}{2} \right)} \right] - \left(\Phi_B + \frac{V_{diel}}{2} \right) \exp \sqrt{\left(\Phi_B + \frac{V_{diel}}{2} \right)} \right\}. \quad (28.10)$$

Here A is a constant, $B = 4\pi/h(2m^*q)^{1/2}$, Φ_B is the potential-energy barrier associated with the tunneling process, t_{diel} is the physical thickness of the dielectric, V_{diel} is the voltage drop across the dielectric, and m^* is the electron effective mass in the dielectric. From (28.10), one observes that the tunneling (leakage) current increases exponentially with decreasing barrier height and thickness for electron direct-tunneling transport. In the context of charge transport through the dielectric, electrically active defects (electron or hole traps) can result in charging of the dielectric, which in turn deleteriously affects the electric field in the channel region and therefore mobility. Conduction mechanisms through such fixed charge traps can also be evaluated through electrical characterization techniques [28.18]. In the case of dielectrics layers, tunneling transport has also been previously examined [28.39].

The introduction of a dielectric material that exhibits a suitable energy barrier and thickness, while performing electrically as a thin SiO₂ layer, is now a key research and development goal for the industry to continue the scaling trend. It is also clear that the gate dielectric must exhibit adequate thickness uniformity and integrity over the surface of the wafer. The reduction of this leakage current impacts important circuit properties such as stand-by power consumption, as shown in Fig. 28.11. With the minimization of leakage current as a key driving force, alternate gate dielectric materials appear to be required for low-power CMOS device technologies in the near future and are therefore the subject of intense research.

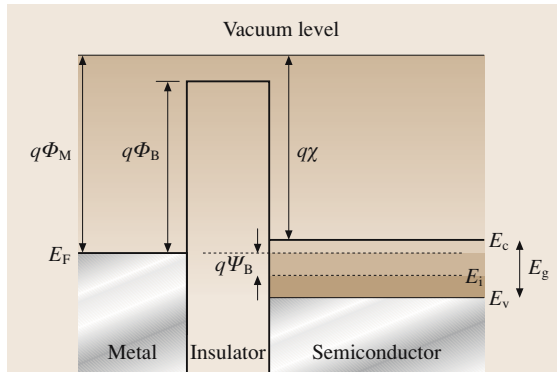


Fig. 28.10 Energy-band diagram for an MIS stack structure

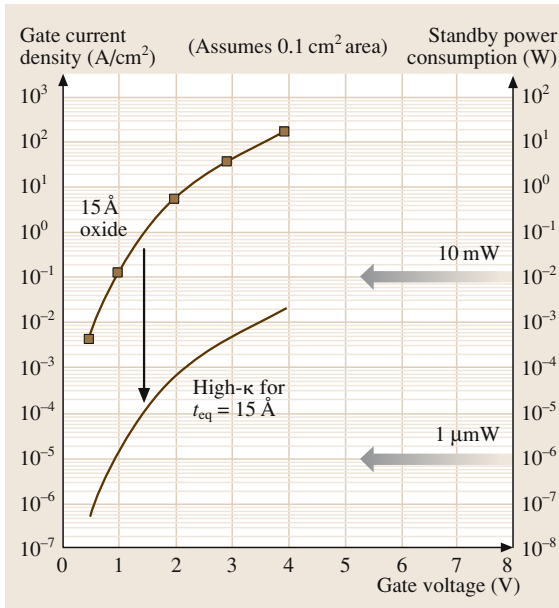


Fig. 28.11 Reduction of leakage (tunneling current) from the incorporation of a higher permittivity gate dielectric. After [28.1] (AIP, with permission © 2001)

Reliability

Any gate dielectric material must possess the ability to enable a specified, stable operational lifetime for the transistor. Key areas of concern in this regard includes defect generation/mitigation resulting in dielectric breakdown phenomenon, charge formation within the dielectric and at the associated interfaces, and resistance to energetic (“hot”) carrier interactions. This is often evaluated in the context of accelerated lifetime testing through electrical stressing in combination with thermal stressing of devices. Reviews on these topics can be found in this handbook as well as other sources [28.6, 28.40, 41].

28.1.3 Silicon Dioxide

As articulated by *Hummel*, silicon has an amazing number of desirable materials (and therefore electronic) properties that no other electronic material can rival [28.42]. Silicon constitutes 28% of the Earth’s crust, it is nontoxic, exhibits a useful band gap ($E_G = 1.12 \text{ eV}$) for applications at room temperature, and can be grown in crystalline form (from abundant sand, i.e. SiO_2). However, it can be argued that the ability of Si to form the stable, high-quality insulator, SiO_2 , may be the primary reason that Si-based transistor technology

Table 28.2 Properties of SiO_2

Geometrical parameters	
Si–Si bond length	3.12 Å
Si–O bond length	1.62 Å
O–O bond length	2.27 Å
Mean bond angle	144° (tetrahedral bonding)
Bond angle range	Si–O–Si: 110° to 180°
Density (g/cm^3)	
Thermally grown (fused silica)	2.20
Quartz	2.65
Index of refraction (optical frequencies)	
Thermally grown (fused silica)	1.460
Quartz	1.544
Quasi-static dielectric constant ($\leq 1 \text{ kHz}$)	
Thermally grown (fused silica)	3.84
Quartz	3.85
Band gap	
Thermally grown (fused silica)	8.9 eV
Quartz	$\approx 9.0 \text{ eV}$
Electrical breakdown strength	
Thermally grown (fused silica)	10 MV/cm
Quartz	$\approx 10 \text{ MV/cm}$

has dominated the industry since the 1960s [28.8, 43]. This compound has a number of material properties that result in outstanding electrical performance. Some of these properties are summarized in Table 28.2. Over the last 25 years, abundant compendia of the properties of SiO_2 with particular emphasis on materials and electrical properties have been published, as well as conferences which have focused almost exclusively on SiO_2 [28.44–53].

It is noteworthy that the early development of transistor technology utilized the semiconductor Ge, which has a rather unstable oxide, GeO_2 , that is water soluble. The utilization of crystalline Si and its stable oxide presented a superior solution for the practical manufacturing of the planar transistor and the associated integrated circuit of the era [28.43]. It was the utilization of the properties of SiO_2 that resulted in the dominance of Si-based device technology over Ge-based technology in the 1960’s.

Bonding Arrangements for Thin SiO_2

As can be seen in Fig. 28.12, the phase diagram for SiO_2 has a number of crystalline phases and associated polymorphisms [28.52, 54]. In the context of accessible phase space appropriate for typical CMOS processes (pressure $\approx 1 \text{ bar}$ and $T \leq 1100^\circ\text{C}$), the quartz and tridymite

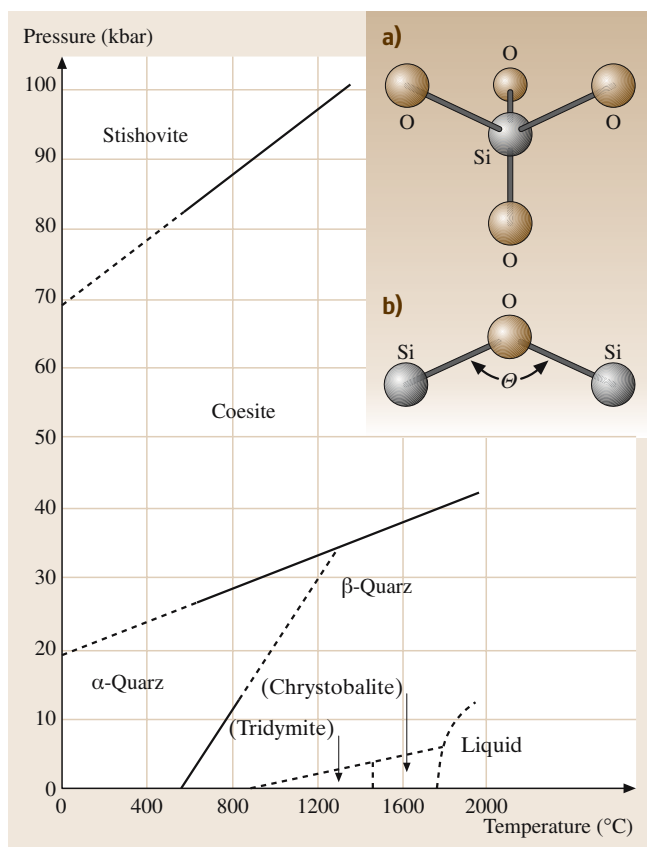


Fig. 28.12a,b Phase diagram for SiO_2 . *Inset:* bonding configurations for the O–Si–O system. **(a)** Tetrahedral unit for the Si–O bonding arrangement. **(b)** Bond angle θ defined for O–Si–O bonds. After [28.54] (IOP Publishing © 1994, with permission)

phases are stable [28.55]. High-pressure phases are also well studied [28.56].

The bulk SiO_2 tetrahedral bonding coordination (inset, Fig. 28.12) results in an average Si–O–Si bond angle of $\theta \approx 145^\circ$ [28.57] comparable to that observed for vitreous silica [28.54, 58]. As noted originally in the classic work by Zachariasen, the bonding interconnections among the various SiO_4 tetrahedra can result in an amorphous SiO_2 film consisting of locally ordered ring structures commonly observed in silica glasses [28.50, 52, 59]. The numerous available variations of the ring structure provide substantial flexibility and minimize bond strain and the density of defects at interfaces [28.60].

For the purposes of microelectronic applications, we first consider so-called *thermally grown* SiO_2 . As the name implies, a SiO_2 layer is formed on Si

from a thermally activated oxidation process (often referred to as a *thermal oxide*) rather than simply a (low-temperature) deposition process. Thermal oxide films are stable enough to endure high-temperature ($\approx 1000^\circ\text{C}$) post-implantation/dopant activation anneals. These films have useful electrical properties including a stable (thermal and electrical) interface with minimal electrically-active-defect densities ($\approx 10^{10}/\text{cm}^2$) as well as high breakdown strength ($\approx 10\text{ MV/cm}$) over a large area ($\approx 1\text{ cm}^2$) [28.18]. The breakdown strength is important for high reliability of devices fabricated with a dielectric, as the scaling of devices typically results in somewhat higher electric fields across the oxide.

Recent experimental results on thin thermal SiO_2 [28.34, 40, 61] combined with modern computational modeling [28.57, 62] of the defects associated with the $\text{SiO}_2/\text{Si}(001)$ interface, which is the dominant microelectronic materials system, and *bulk* SiO_2 [28.63] has provided a reasonably consistent picture of the known defects and interface structure.

High-resolution X-ray photoelectron spectroscopy (XPS) evidence, coupled with other experimental and theoretical modeling results, appears to support the presence of SiO_x ($x = 4$) species (often called *sub-oxides*) detected as a chemical shift in the associated Si oxidation states (Fig. 28.13) [28.64–67].

More recent high-resolution angle-resolved XPS analysis of ultrathin (6-Å) SiO_2 has resulted in a further understanding of the average oxidation state observed with depth along the interfacial transition region [28.68, 69]. From these experimental results, models of the interface have been constructed from first-principles methods that reproduce the experimental results, as shown in Fig. 28.14 [28.57, 62]. As seen in Fig. 28.14, the transition region between the Si substrate and the SiO_2 consists of interfacial suboxide species of varying abundance (and therefore density). These results appear to be in general agreement with recent high-resolution transmission electron microscopy as well [28.70].

Oxidation Kinetics of Si

The oxidation kinetics of Si have also been exhaustively studied and reviewed [28.28, 40, 71]. Two regimes for oxidation are generally characterized as *passive* and *active* in nature. The passive oxidation regime refers to the low-temperature/high-pressure region of the pressure–temperature (p – T) phase space shown in Fig. 28.15 [28.40, 72–77]. In this portion of the p – T phase space, the reaction between Si and O_2 proceeds

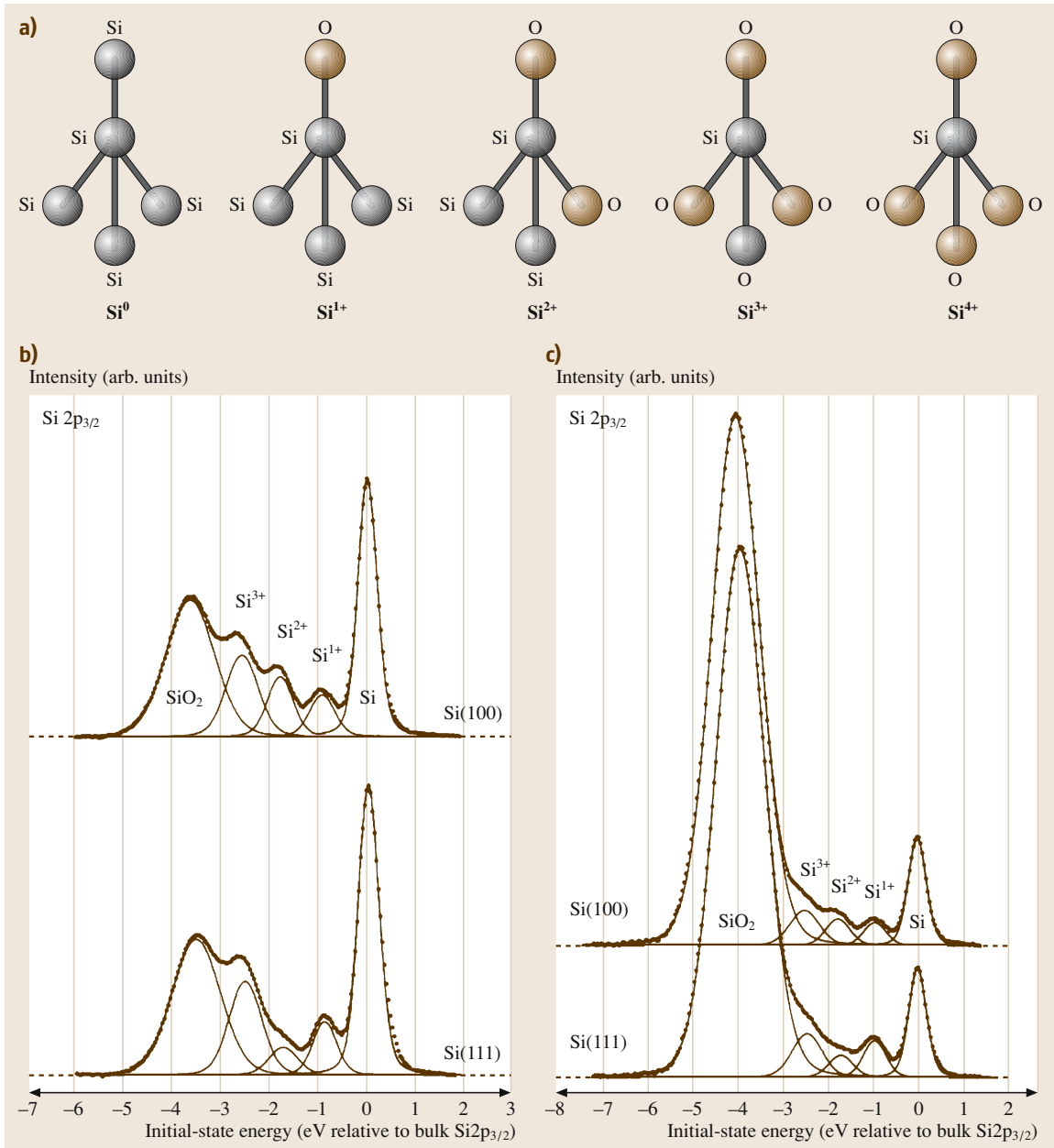


Fig. 28.13 (a) Oxidation states and local chemical bonding variations associated with the Si–SiO₂ interface. After [28.54]. High-resolution XPS results demonstrating the existence of such bonding variations at the interface for substrate orientation and with (b) 5-Å and (c) 14-Å SiO₂ film thickness. Photon energy $h\nu = 130$ eV. After [28.66] (APS, with permission © 1988)

as a surface reaction resulting in a thin (≤ 2 -nm) SiO_x layer, where $0.5 \leq x \leq 2$. Decreasing pressure and increasing temperature during the reaction re-

sults in both Si etching and oxidation – the so-called *active* regime – and the formation of thicker SiO₂ films.

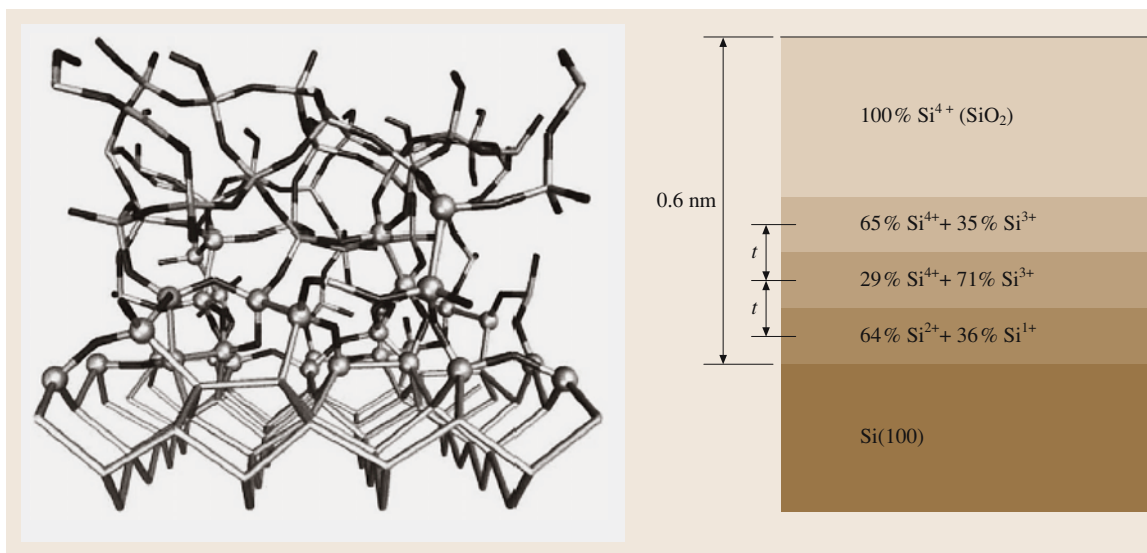


Fig. 28.14 (a) First-principles simulations of the SiO₂/Si interface. After [28.62] © 2002 Elsevier. (b) Angle-resolved XPS results for the SiO₂/Si interface showing the detected O–Si–O bonding environments. The single layer thickness $t = 0.137$ nm associated with the bonding distance in Si in this evaluation. After [28.69] (© 2001 by APS, with permission)

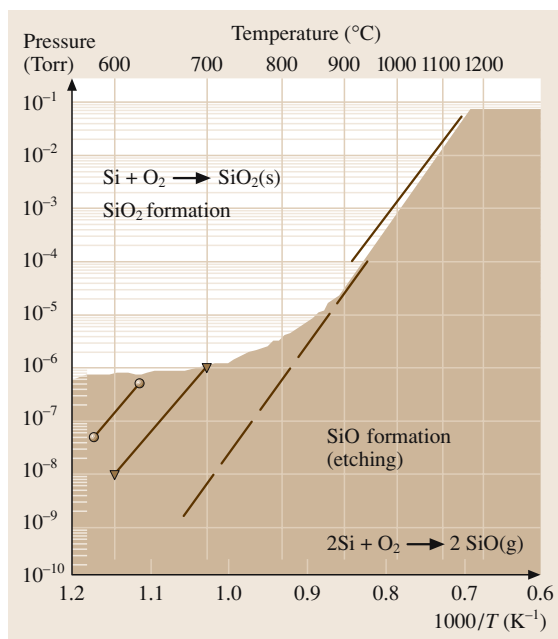


Fig. 28.15 Pressure–temperature phase diagram for thin SiO₂/Si. After [28.77] (© 1997 AIP, with permission)

The clever utilization of isotopic labeling techniques coupled with depth profiling from ion scattering [28.78]

and nuclear reaction analysis [28.71] has generally confirmed that molecular oxygen (O₂) is the mobile species that interstitially diffuses through a growing oxide and only reacts at the Si interface to form the SiO₂ on the Si surface, as first empirically proposed by *Deal and Grove* in 1965 for SiO₂ thicker than about 4 nm [28.79,80]. The details of the oxidation process for thin films deviates from the Deal–Grove model, and is discussed thoroughly in the review by *Green et al.*, for example [28.40].

Defects for Thin SiO₂

In addition to the electrical breakdown strength associated with SiO₂, the physicochemical properties of the interface between the substrate Si and SiO₂ plays a major role in establishing the observed electrical properties [28.61,67]. Like all interfaces of solids, the interface between bulk Si and bulk SiO₂ naturally results in the presence of point defects (Fig. 28.16). Examples include strained bonds, atoms that are not fully bonded to neighboring atoms (so-called “dangling”), bonds and hydrogenic species bonded among the Si–O bonding network. These defects can result in energy levels in the band structure associated with an MIS device, and serve as *traps* for carriers (electrons and holes).

The dominant orientation for contemporary Si-based integrated circuit technology substrates is Si(100). This is, of course, no accident and the choice of the substrate

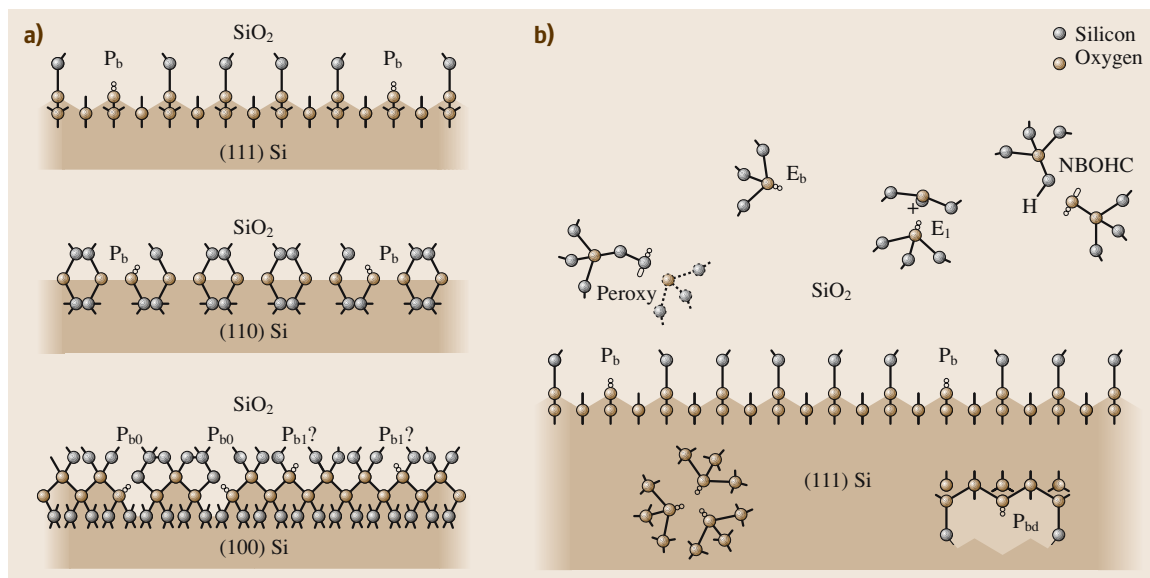


Fig. 28.16a,b Schematic representations of **(a)** the various point defects for the SiO_2/Si interface and **(b)** bulk SiO_2 . After [28.54] (IOP, with permission © 1994)

orientation has much to do with the resultant SiO_2/Si interface. Early in the development of Si-based transistors, it was observed that the density of interface states for SiO_2 grown on Si(100) was significantly lower than that observed on Si(111) [28.81–85]. This difference in interface state behavior is attributed to the density of unsatisfied (*dangling*) bonds at the SiO_2/Si interface.

Techniques sensitive to the presence of such point defects, such as electron (paramagnetic) spin resonance (EPR or ESR) [28.88] have provided valuable insight on the atomic nature of the various defects in the bulk and interfacial regions of SiO_2 [28.89–94]. For such defects, although the electrical manifestation of significant densities ($\geq 10^{10}/\text{cm}^2\text{eV}$) of such *trap states* is easily detected by conventional electrical characterization methods, the unambiguous identification of the moiety responsible for the observed electrical behavior is very difficult using only electrical characterization. From techniques like ESR employed over the last 30 years, point-defect structural models have been established in the bulk regions of SiO_2 and Si, as well as at the interface (including various substrate orientations). A schematic illustration of various known point defects for the Si/ SiO_2 system is shown in Fig. 28.16. The major classes of defects that have been examined over the last 35 years by ESR have included interfacial dangling bonds (P_b centers) and bulk defects in SiO_2 (E' centers) [28.91, 94].

The early discovery and study of P_b centers [28.95–97] eventually led to the interpretation of their ESR signals as due to dangling bonds and the realization of their importance to MOS device physics [28.86, 90,

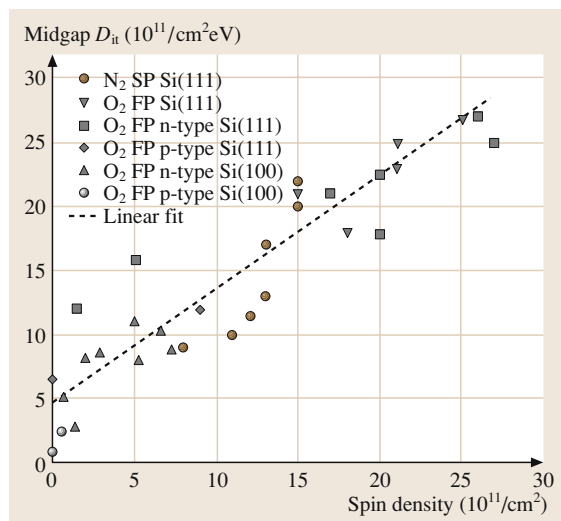


Fig. 28.17 Correlation of mid-gap interface state density with detected spin density from ESR measurements for $\text{SiO}_2/\text{Si}(111)$ and $\text{SiO}_2/\text{Si}(100)$. FP and SP denote *fast* and *slow* pullout conditions for the wafers after thermal treatments. After [28.86, 87], with permission

91, 98]. It has also been noted that P_b -center defects are generated for unannealed, thermally stressed and radiation-damaged MOS structures, and moreover can account for roughly 50% of the density of interface defect states (D_{it}) for Si/SiO₂ interfaces [28.54, 99, 100]. It has been demonstrated that the interface state density can be directly proportional to the density of P_b centers (dangling bonds) for the SiO₂/Si(111) and SiO₂/Si(100) interfaces, as seen in Fig. 28.17. Recently, further ESR work has been performed to establish the detailed structure of the defect on Si(100) and Si(111) [28.101–108].

Mitigation of Defects for Thin SiO₂

Synopses of the early MOS transistor work [28.109–111] examining the importance of the dielectric–semiconductor interface are available [28.6, 43, 112]. Indeed, the close connection between interfacial chemical behavior and electrical device performance was investigated and realized in pioneering surface science work by Law and coworkers on the reaction of gaseous species with atomically clean Ge [28.113] and Si surfaces [28.114–116]. In particular, the relative interface state (dangling-bond) density, as measured by the areal density of surface reactions with technologically important species such as H₂O, H₂, O₂, CO, CO₂, provided important clues on the control of the Si/SiO₂ interface and the resultant electrical properties reported 5–10 years later.

The control of the density of interface defects through the chemical reaction of species, introduced mainly through gaseous exposure at elevated temperatures, has proven to be fruitful. As noted above, early surface science work [28.114] indicated the rapid reaction of H₂ with the atomically clean Si surface. By definition, the atomically clean Si surface is saturated with dangling bonds. The reaction of these bonds with H₂ results in the chemical passivation of the surface – that is, the reaction of the surface to eliminate the dangling bonds and produce a H-terminated Si surface. In the context of the MOS structure [28.47, 117], it was realized early that annealing the structure in ambient H₂ resulted in beneficial transconductance performance [28.118–121]. The use of anneals in forming gas (N₂:H₂ of various mixtures, typically 90–95%N₂:10–5%H₂) were originally developed to improve electrical contacts for the gate and source/drain regions of the MOSFET. Balk and Kooi established the effect of hydrogen ambients on the reduction of fixed charge in the Al/SiO₂/Si MOS structure [28.120, 122].

Subsequent studies demonstrated that anneals of the MOS structure in hydrogenic environments, typically at

400–500 °C for 30–60 min, results in the passivation of dangling bonds at the interface [28.86, 101–106, 123–126].

Hydrogen incorporation into the bulk of SiO₂ can, however, also be detrimental to dielectric performance in MOS capacitors and FETs [28.54, 127]. For example, silicon bonded to hydroxyl (silanol) species have been identified with fixed charge in the oxide, resulting in undesirable, irreversible voltage shifts. This charge induced shift is shown in Fig. 28.18 for n-type and p-type MIS diode structures from their associated capacitance–voltage response [28.1, 6, 7]. (An analogous threshold-voltage shift would be observed in a transistor turn-on characteristic.) More complicated effects such as negative-bias temperature instability (NBTI), where an increase of the density of fixed charge (Q_f) and interface trap (Q_{it}) density is noted with time upon thermal stress and/or under negative bias, has been attributed to H₂O-induced depassivation of Si dangling bonds (i. e., generation of P_b centers) at the Si/SiO₂ interface [28.54, 128]. Defects generated by radiation damage have also been extensively investigated [28.54, 129] as well as interactions with annealing ambients such as vacuum [28.130] or SiO [28.124].

Dielectric Breakdown and Reliability of SiO₂

As noted in Sect. 28.1.2, the reliability of dielectrics is obviously an important phenomenon to control. The scaling of microelectronic devices necessarily results in increased stress on the dielectric due to the higher electric fields placed across the dielectric film. As a result, power-supply voltages are also scaled to minimize the likelihood of catastrophic (*hard*) breakdown, which would generally result in complete failure in the associated integrated circuit. As noted by Hori, such a catastrophic breakdown phenomenon is dependent upon the presence of defects in the dielectric, and thus requires a statistical analysis of many devices (and therefore films) to enable a reliability prediction for the dielectric layer [28.18]. Evaluation of the breakdown is often performed under the conditions of constant-field stress until a time at which breakdown is observed [called time-dependent dielectric breakdown (TDDB)]. Breakdown fields for thermally grown SiO₂ with thicknesses larger than ≈ 10 nm often exceed 10 MV/cm, providing outstanding insulator prosperities for microelectronic applications. Various models have also been developed to accelerate such testing to perform reliability predictions. A concise review of these is offered in the literature as well [28.18, 28]. Defects such as film nonuniformity, bond stress, surface asperities, contam-

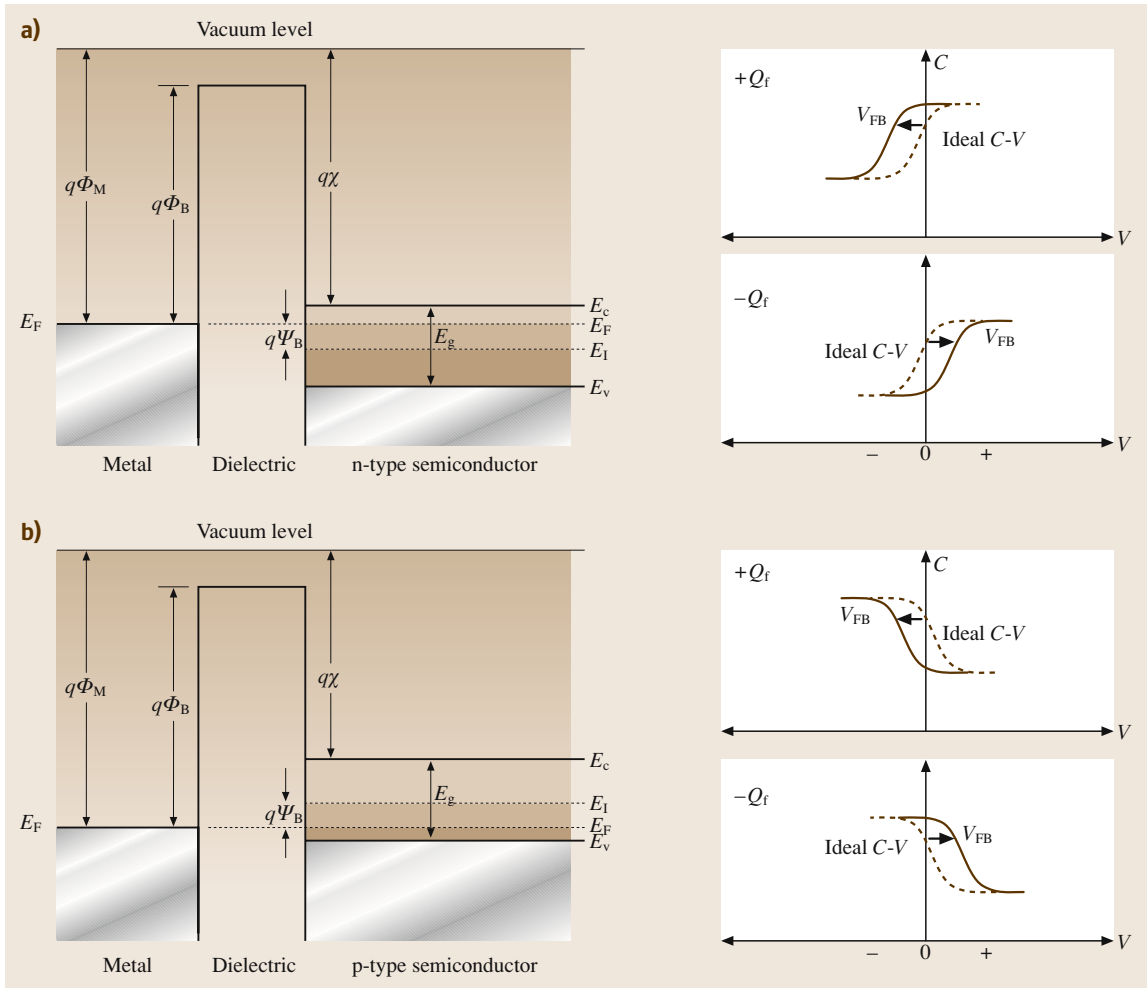


Fig. 28.18a,b The effect of fixed charging of the capacitor dielectric, demonstrating the shift of the associated flat-band voltage for MIS structures with (a) n-type and (b) p-type substrates. After [28.1] (© 2001 AIP, with permission)

ination, and particulates embedded in the film or the Si substrate are a few examples of potential causes of catastrophic breakdown phenomenon.

With the scaling of the SiO_2 gate dielectric layer to thicknesses well below 10 nm, extensive evaluations have been made of the various models that attempt to predict the reliability of SiO_2 in regard to breakdown. Often, a *soft* breakdown can be observed, where the film exhibits a sudden increase in conductivity, but not to the same degree as that observed in hard breakdown phenomenon. In spite of the many years of research on SiO_2 thin films, the nuances of the dependence of voltage acceleration extrapolation on dielectric thickness and the improvement of reliability projection arising

from improved oxide thickness uniformity, have only recently become understood, despite decades of research on SiO_2 [28.131].

28.1.4 Silicon Oxynitride: SiO_xN_y

The introduction of nitrogen in SiO_2 , often described as *silicon-oxynitride* (or more simply SiON) has provided the opportunity to maintain the scaling expectations for integrated circuits while minimizing the process variations associated with the modification of the gate dielectric material [28.18, 40, 71, 132]. (It should be noted that engineers and technologists often loosely refer to SiON films simply as oxide or nitride – frequently lead-

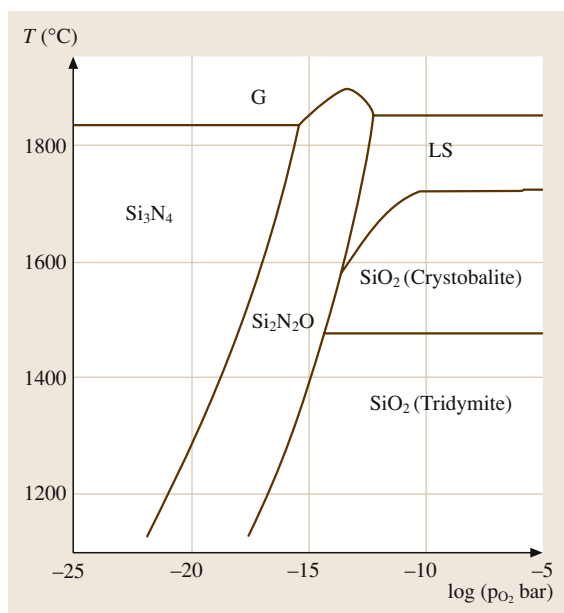


Fig. 28.19 Calculated phase diagram for the growth of Si—O—N for various oxygen partial pressures. After [28.134], with permission

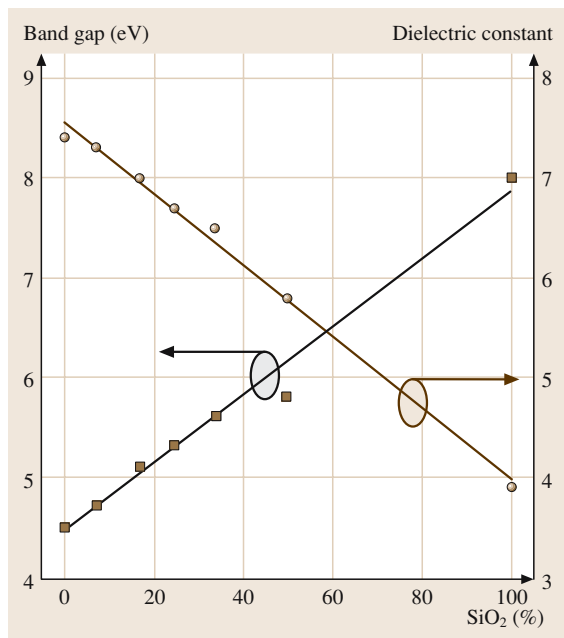


Fig. 28.20 Dependence of permittivity and insulator band gap of SiO_xN_y with SiO_2 content. After [28.135], with permission

ing to confusion on the detailed chemical composition of the film.) Introduction of N into SiO_2 has been accomplished by thermal treatments (in N_2 , N_2O , NO and NH_3) as well as plasma treatments [28.18, 133]. The bulk phase diagram for the SiON system is presented in Fig. 28.19, which indicates that, under equilibrium conditions, $\text{Si}_2\text{N}_2\text{O}$ is the only stable SiON species and SiO_2 and Si_3N_4 would not coexist [28.134]. As noted by Green et al., N would not be expected to incorporate into SiO_2 whenever an even very small partial pressure of oxygen $P_{\text{O}_2} > 10^{-20}$ atm is present. However, the nonequilibrium surface reaction kinetics and/or a reduction in interfacial bond strain likely plays a key role in the tendency for interfacial N incorporation for SiON films [28.40].

The dielectric constant and band gap as a function of the SiO_2 content are presented in Fig. 28.20 [28.135]. As noted in Sect. 28.1.2, it is the increase in the dielectric constant that increases the capacitance of an associated MOS structure. Thus, for a SiON film with $\kappa = 7$, a 2-nm SiON film would ideally exhibit electrical behavior (such as capacitance and leakage-current behavior) equivalent to a 1.1-nm SiO_2 film, according to (28.8). The data in Fig. 28.20 implies that the index of refraction (at 635 nm) varies between $n(\text{SiO}_2) = 1.46$ and $n(\text{Si}_3\text{N}_4) = 2$, indicating that optical techniques such as ellipsometry can provide very useful information for such films.

Like the SiO_2/Si system, point defects have also been studied for SiON films on silicon [28.136]. Lenahan and coworkers examined the role of strain at the interface and the impact on the associated dangling bonds [28.137]. More recently, work on the effects of hydrogen annealing has also examined the SiON system [28.138].

It is also noted that the SiON material system also exhibits useful diffusion-barrier properties. This property is utilized to inhibit uncontrolled dopant diffusion from the polysilicon gate electrode through the dielectric layer and into the channel region of transistors, for example. The formation of SiON films from the reaction of NO or N_2O with Si, as well as nitridation of Si and SiO_2 from N_2 and NH_3 exposure, has been examined in detail and has been summarized in recent reviews [28.40, 71].

For transistor gate dielectric applications, a controlled variation in the depth concentration profile for N is often required throughout the dielectric to maintain the improved device reliability and mobility associated with SiON thin films [28.28]. An example of such N profiles is shown in Fig. 28.21, where a somewhat reduced N content is noted in the vicinity of the Si channel/dielectric interface relative to that near the dielectric/poly-Si gate

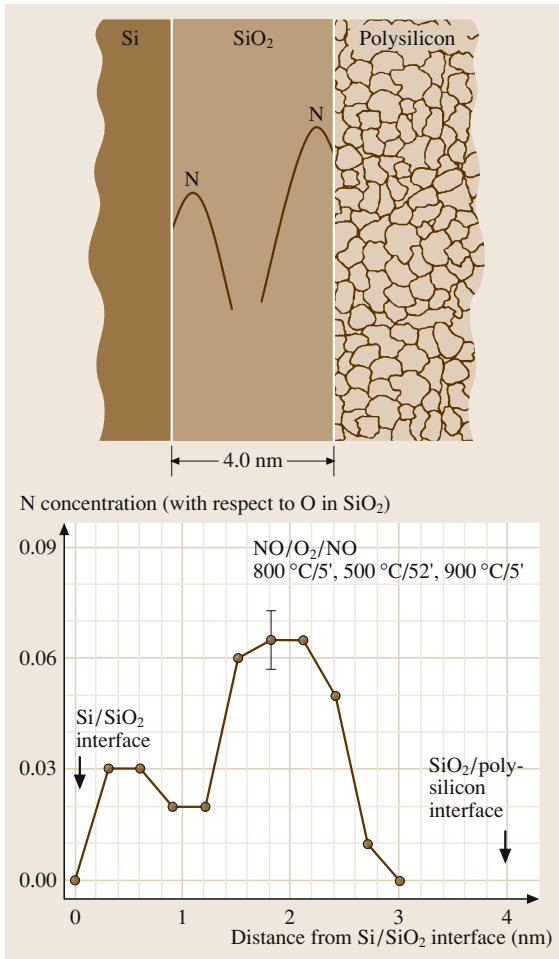


Fig. 28.21 N concentration depth profiles for tailoring the overall dielectric constant and materials properties of a SiON layer. After [28.34] (© 1999 Elsevier, with permission)

interface [28.34]. Such profiles are thought to control effects such as interfacial strain and roughness (which can degrade mobility) and inhibit dopant out-diffusion from the gate electrode. Incorporation of N in thin dielectric films using plasma processes has been studied extensively [28.133, 139–143].

Given the various leakage-current constraints afforded by power limitations and reliability, the scaling of SiON appears to be limited to $t_{eq} \approx 1$ nm, corresponding to the 65-nm node for high-performance microprocessor products set to begin in ≈ 2007 [28.12]. As a result, investigations of alternative high- κ dielectric materials have been initiated over the last several years.

28.1.5 High- κ Dielectrics

Recent research has focused on high- κ dielectrics to further enable scaling of transistor (and memory capacitor) technology. In the context of the industry roadmap, the term high- κ dielectric more generally refers to materials which exhibit dielectric constants higher than the SiON films ($\kappa \approx 7$) described above.

Desirable Properties for High- κ Dielectrics

Many materials systems are currently under consideration as potential replacements for SiO₂ and SiO_xN_y as the gate dielectric material for sub-100-nm (CMOS) technology. It should be stated at the outset that this field of materials research remains very active, and a conclusive summary on the topic of high- κ materials, particularly for conventional MOSFET gate dielectric applications, is not yet possible. The field continues to evolve with interesting research being reported almost weekly in peer-review journals as well as industry newsletters and magazines. Nevertheless, a systematic consideration of the required properties of gate dielectrics indicates that the key desirable properties for selecting an alternative gate dielectric include: permittivity and the associated band gap/alignment to silicon as well as mobility, thermodynamic stability, film morphology, interface quality, compatibility with the current or expected materials to be used in processing for CMOS devices, process compatibility, and reliability. The desirable properties are summarized in Table 28.3. Reviews of the developing field and recent research on high- κ dielectric materials candidates are available [28.1, 2, 144].

Many gate dielectric materials appear favorable in some of these areas, but very few materials are promising with respect to all of these properties. Indeed, many of these desirable properties are interrelated. Many of these materials have been examined over the last 20 years for capacitor applications, for example. Materials currently under extensive investigation include oxides, silicates and aluminates of Hf, Zr, La, Y, and their mixtures, which results in dielectric constants in the range 10–80. Table 28.4 summarizes some of the measured or calculated properties associated with high- κ dielectric material constituents. There are also some recent studies of oxides associated with the rare-earth lanthanide series of elements in the periodic table [28.145]. One notes the tradeoff between the band gap (and therefore the conduction/valence-band offsets to those of Si) for these dielectrics and the associated dielectric constant, as seen in Fig. 28.22 [28.146].

Table 28.3 Desirable properties for high-κ gate dielectrics (65-nm node)

Physical property	Value/Criteria	Electrical property	Value/Criteria
Permittivity	15–25	Equivalent oxide thickness	< 1 nm
Band gap	> 5 eV	Gate leakage current (low power)	< 0.03 A/cm ²
Band offset	> 1.5 eV	Gate leakage current (high performance)	< 10 ³ A/cm ²
Thermodynamic stability to 1000 °C	When in direct contact with Si channel	CV dispersion	Minimal (meV)
Compatibility with metal electrodes	Mimimized extrinsic (pinning) defects	CV hysteresis	Minimal (meV)
Morphology control	Resistance to interdiffusion of constituents, dopants, and capping metals	V _T (V _{FB}) shift (fixed charge, defects, trapping, etc.)	Minimal
Deposition process	Suitable for high-volume production	Channel mobility	Near SiO ₂ universal curve
Etching	Suitable control for patterning after processing/annealing	Interface quality	Near SiO ₂ ; D _{IT} ≈ 5 × 10 ¹⁰ /cm ² eV

Table 28.4 Comparison of relevant properties for selected high-κ candidates. Key: mono. = monoclinic; tetrag. = tetragonal

Material	Dielectric constant (κ)	Band gap E _G (eV)	ΔE _C (eV) to Si	Crystal structure(s) (400–1050 °C)
SiO ₂	3.9	8.9–9.0	3.2–3.5 ^b	amorphous
Si ₃ N ₄	7	4.8 ^a –5.3	2.4 ^b	amorphous
Al ₂ O ₃	9	6.7 ^h –8.7	2.1 ^a –2.8 ^b	amorphous*
Y ₂ O ₃	11 ^d –15	5.6–6.1 ^d	2.3 ^b	cubic
Sc ₂ O ₃	13 ^d	6.0 ^d		cubic
ZrO ₂	22 ^d	5.5 ^a –5.8 ^d	1.2 ^a –1.4 ^b	mono., tetrag., cubic
HfO ₂	22 ^d	5.5 ^d –6.0	1.5 ^b –1.9 ^c	mono., tetrag., cubic
La ₂ O ₃	30	6.0	2.3 ^b	hexagonal, cubic
Ta ₂ O ₅	26	4.6 ^a	0.3 ^{a,b}	orthorhombic
TiO ₂	80	3.05–3.3	≈ 0.05 ^b	tetrag. (rutile, anatase)
ZrSiO ₄	12 ^d	6 ^d –6.5	1.5 ^b	tetrag.
HfSiO ₄	12	6.5	1.5 ^b	tetrag.
YAlO ₃	16–17 ^d	7.5 ^d		**
HfAlO ₃	10 ^e –18 ^g	5.5–6.4 ^f	2–2.3 ^f	**
LaAlO ₃	25 ^d	5.7 ^d		**
SrZrO ₃	30 ^d	5.5 ^d		**
HfSiON	12–17 ^{i,j}	6.9 ^k	2.9 ^k	amorphous

* (γ-Al₂O₃ phase) has been recently reported [28.148], ** Onset of crystallization depends upon Al content, ^a [28.149], ^b [28.150–152], ^c [28.153], ^d [28.146], ^e [28.154], ^f [28.153], ^g [28.155], ^h [28.156], ⁱ [28.157], ^j [28.158], ^k [28.159]

The compatibility of alternative dielectrics with metal gate electrodes is also an important consideration [28.1, 3, 147]. The interfacial reactions between some gate electrode metals and gate dielectric is thought to lead to extrinsic states due to Fermi-level pinning, which shifts the threshold voltages for transistors to

a fixed value. Another clear challenge is compatibility with other process steps that entail substantial thermal budgets (≈ 1000 °C, ≤ 10 s) for conventional CMOS. Future scaling may well require a modification (reduction) of such thermal budget desires to incorporate metal gate electrodes in CMOS, and thus may open

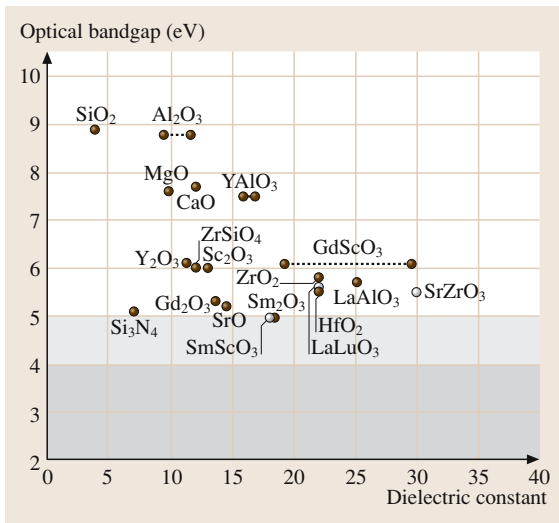


Fig. 28.22 Band gap versus dielectric constant for a number of dielectric materials. After [28.146] MRS Bulletin, with permission

the door to the consideration of other gate dielectric materials that exhibit stability at somewhat lower temperatures. However, key issues that must be addressed under such a scenario include alternative source/drain engineering (e.g. dopant activation at lower thermal budgets) and/or gate electrode formation in the device fabrication flow (e.g. gate electrode insertion after high-temperature anneals). Previous research on such CMOS process modifications has indicated that a variety of challenges exist to address such scenarios with adequate process margin and yield.

Mobility Degradation for Transistors with High- κ Dielectrics

In the case of many of the alternate gate dielectric materials (mainly metal oxides) currently under consideration, the polarizability of the metal–oxygen (*ionic*) bond is responsible for the observed low-frequency permittivity enhancement. Such highly polarizable bonds are described to be *soft* relative to the less polarizable *stiff* Si–O bonds associated with SiO₂. Unlike the relatively stiff Si–O bond, the polarization frequency dependence of the M–O bond is predicted to result in an enhanced scattering coupling strength for electrons with the associated low-energy and surface optical phonons. This scattering mechanism can therefore degrade the electron mobility in the inversion layer associated with MOSFET devices. A theoretical examination of this effect is provided by Fischetti et al. where calculations

of the magnitude of the effect indicate that pure metal–oxide systems, such as ZrO₂ and HfO₂, suffer the worst degradation, whereas materials which incorporate Si–O bonds, such as silicates, fare better [28.160]. In that work, it is also noted that the presence of a thin SiO₂ interfacial layer between the Si substrate and the high- κ dielectric can help boost the resultant mobility by screening this effect, although the maximum attainable effective mobility is still below that for the ideal SiO₂/Si system. These researchers further suggest that the effect is also minimized by the incorporation of Si–O in the dielectric, as in the case of pseudo-binary systems such as silicates.

Comparisons of this model with experimental evidence have recently been reported explicitly through the comparison of HfO₂ and Hf–silicate mobility studies. The results indicate that n-MOS and p-MOS poly-Si gated devices with Hf–silicate dielectrics exhibit better mobilities (approaching those of SiO₂) over those obtained using HfO₂ due to a diminished soft-phonon scattering mechanism [28.161]. Incorporation of metal gate electrodes results in further improvement [28.162].

Current High- κ Dielectric Research and Development

At this time (end of 2004), work on Hf-based dielectric materials, mainly Hf–silicates [28.163] and HfO₂ [28.1], dominates the recent engineering literature (Fig. 28.23) and they appear to exhibit useful properties for integrated-circuit scaling down to the 20-nm node. These properties include a relative stability for interfacial reactions that prefer silicate formation thus can avoid a lower overall dielectric constant of the high- κ dielectric stack, due to a pure SiO₂ interfacial layer, as described by equation (28.9). The papers on these materials systems are too numerous to list individually here and so the reader is referred to the reviews on the topic [28.1, 2, 144].

Of particular importance for gate dielectrics have been investigations of the stability (in particular, changes in morphology and interdiffusion) of these films upon thermal processing in view of the required integration constraints for CMOS applications [28.3]. A significant hurdle for the integration of all dielectrics currently under consideration for conventional CMOS is the stability of all gate-stack constituents (stable film morphology, minimal interdiffusion, etc.) under a thermal budget of $\approx 1000^\circ\text{C}$ for 10 s for dopant activation and adequate process integration margin. Under such thermal treatments, gate dielectric constituents such as Zr [28.164] or Al [28.165] have been reported to penetrate the Si chan-

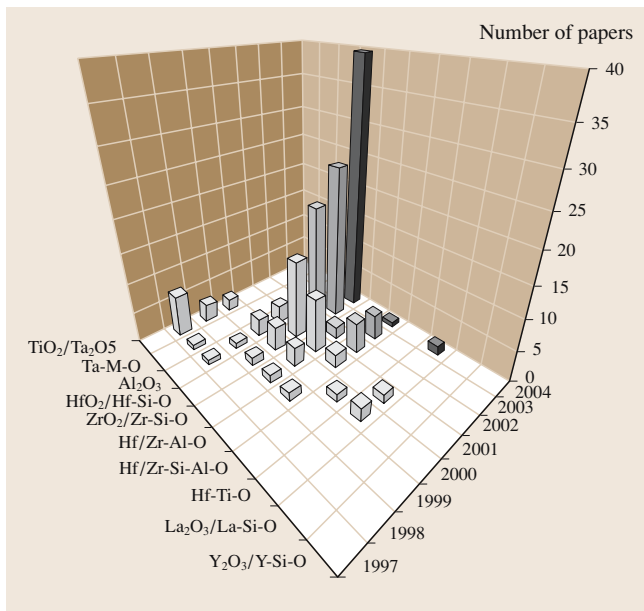


Fig. 28.23 Distribution of papers presented at recent IEEE international electron devices meeting (IEDM) and very-large-scale integration (VLSI) conferences indicate the current emphasis of materials systems examined for gate dielectric research

nel region and thus present a potential source of impurity scattering for mobility degradation. In contrast, Hf does not appear to exhibit such penetration within detectable limits [28.166].

It is also important to note that most of the high- κ metal oxides listed in Table 28.4 crystallize at relatively low temperatures ($T \approx 500\text{--}600^\circ\text{C}$). An exception is Al_2O_3 where crystallization in thin films is observed at $T \approx 800^\circ\text{C}$ [28.167]. In the case of Hf-silicates, film morphology (viz. suppression of crystallization) and diffusion-barrier properties can be controlled through the incorporation of N, as in the case of SiO_2 [28.157, 168–172]. Currently, HfSiON dielectrics appear to provide desirable properties in this regard.

Suppression of crystallization is also observed for aluminates (depending upon the Al content) [28.173], but there are fewer studies of interdiffusion with the Si substrate available at this time. As noted above however for the case of Al_2O_3 thin films, Al penetration into the Si substrate has been reported upon thermal annealing budgets appropriate for dopant activation [28.165]. Such

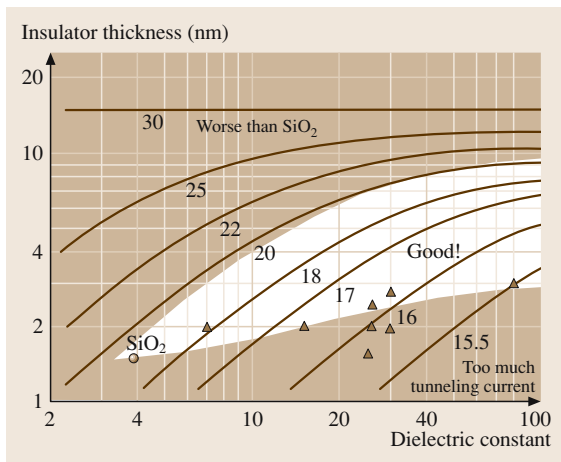


Fig. 28.24 Contours of constant scale length, which is proportional to the transistor gate length, versus dielectric constant and insulator thickness, showing the useful design space for high- κ gate dielectrics. Data points are rough estimates of the tunneling constraints for various high- κ insulators. The depletion depth is assumed to be 15 nm. The useful design space will shrink with decreasing depletion depth. After [28.23] (© 2004 IEEE, with permission)

penetration could be a concern for any high- κ dielectric incorporating Al as well.

The morphology of the dielectric film may also impact on the propensity for capping layer constituents (from metal gate material layers) to diffuse through the stack as well. Metal gate materials currently under investigation include well-known minority-lifetime killers such as Ni (in NiSi) as well as tunable-work-function alloys and layers [28.3, 147, 174, 175]. The potential for the interdiffusion of metals into the channel region is clearly undesirable.

The limits of a useful high- κ dielectric constant value have also been examined. For example, the study by Frank et al. (Fig. 28.24) suggests that there exist a range of useful permittivity values as planar CMOS transistor structures are scaled [28.23]. It is noted that the physical thickness of the gate insulator should be less than the Si depletion length under the channel region, and that this results in a limited design space that enables further device scaling. Again, one notes that the dielectric constant cannot be arbitrarily increased without careful consideration of the complete transistor design.

28.2 Isolation Dielectrics

Dielectrics are also employed to electrically isolate various regions of an integrated-circuit technology. The dominant material for these applications is SiO_2 . A method commonly employed to obtain such isolation in larger-scale IC technologies (generally with gate lengths larger than $0.25\text{ }\mu\text{m}$) is the so-called local oxidation of silicon (LOCOS) technique, where regions of Si between various components are preferentially (thermally) oxidized [28.18, 28]. Such isolation oxides are typically several hundred nanometers thick and the tapered shape of the edge of the LOCOS isolation oxide near the transistor gate dielectric and source/drain region (often referred to as the *bird's beak*) is an important region to control during the device fabrication at these gate lengths. Thinning of the dielectrics in this region results in breakdown reliability concerns, as the electric fields in this region can be quite high. A *reoxidation process* is often performed to improve the thickness and reliability of the SiO_2 in this region as well.

Scaling ICs beyond this gate-length regime however has required the placement of isolated regions utilizing deposited SiO_2 trench structures, as seen in Fig. 28.6. As summarized by Wolf, extensive work has been done to control the shape of the trench walls and the SiO_2 layer initially formed on these walls. Chemical vapor deposition methods are normally utilized for the dielectric deposition as this approach provides superior conformality in the trench structure. Filling the trench without void formation, controlling film stress, and chemical mechan-

ical polishing properties are also important aspects that must be addressed in the fabrication process [28.12, 28]. Scaling of integrate circuits will result in a decrease of the area available for these isolation trenches, and so the increasing aspect ratio of the trench depth to width requires considerable attention in regard to trench filling. The control of the shape of the top corner regions of the trench structures is also an area of concern due to high-field reliability as well as the formation of essentially a parasitic *edge transistor*. Initially, thermal oxidation methods were employed to round off the shape of the corner, but further scaling will require etch processing methods.

The fabrication of transistor source/drain regions utilizing implantation processes also utilizes a dielectric spacer layer that surrounds the transistor gate region (Fig. 28.6). This spacer provides isolation between the gate and source/drain regions, and also permits the control of the depth (and therefore profile) of the implanted dopant species – a so-called self-aligned dopant implantation process. Dielectrics used for spacer technologies include deposited SiO_2 and Si_3N_4 , and the extent of the spacer dimensions from the gate is an important device parameter to control [28.18]. Further scaling of transistor structures will likely result in the need for elevated source/drain regions, and so process compatibility of the spacer material with the source/drain formation processes and high- κ dielectrics will become an important consideration [28.12].

28.3 Capacitor Dielectrics

Capacitors are employed in a variety of integrated circuits including storage (memory) circuit elements and input/output coupling circuitry. Clearly, dielectrics are critically important in this application. In contrast to the MIS structure associated with transistors, capacitors are passive devices incorporating a metal–insulator–metal (MIM) structure. Early electrodes were composed of degenerately doped Si while more recent work focuses on integrating metals for modern devices.

28.3.1 Types of IC Memory

A dominant memory technology for the IC industry includes dynamic random-access memory (DRAM) in which capacitors play the essential role of storing charge,

and thereby useful information. This type of memory element requires refreshing in order to maintain useful information, and the rate of refresh is fundamentally related to the dielectric associated with the capacitor structure and leakage of charge from that capacitor. As a result, this class of memory devices is called volatile. In contrast, a static random-access memory (SRAM) nonvolatile memory element required a capacitor which stores charge without the refresh requirement, and can preserve charge for many years. Again capacitor design must include the consideration of the dielectric [28.27].

In addition to DRAM capacitors, other types of capacitors are used in the back end of the transistor flow. These devices are MIM capacitors, and they typically reside between levels of metal interconnects, and

serve as decoupling or radio-frequency (RF) capacitors for input/output functions. For these applications, the capacitance density is still of critical importance, but other factors must be considered, such as the linearity of the capacitance as a function of voltage and temperature.

28.3.2 Capacitor Dielectric Requirements in View of Scaling

In brief, scaling the capacitor dimensions requires a tradeoff between the amount of stored charge required for a reliable memory element and the area occupied by the capacitor (and the associated transistors for the memory cell). From generation to generation of devices, the amount of stored charge is kept roughly constant (≈ 30 fF/cell), even though dimensional scaling occurs, for bit detection (sense amplifier), retention and reliability reasons [28.28, 176].

As noted in (28.6) for a simple parallel-plate capacitor, the area associated with the capacitor as well as the dielectric permittivity play an important role. Maintaining the required capacitance with scaling was (and will be) accomplished by increasing the capacitor area, for a suitable dielectric thickness where leakage currents are severely limited; this can be accomplished by utilizing three-dimensional capacitor structures produced in the Si bulk, such as deep-trench capacitors [28.176, 177]. Clever and intricate processing methods to increase the capacitor area were adopted for several generations for well-established SiO_2 and Si–oxynitride dielectric capacitor materials, such as the implementation of hemispherical-grain (HSG) poly-Si and *crown* structures (Fig. 28.25). Eventually, practical

concerns about cost and device yield make manufacturing such structures problematic. Thus, substantially increasing the dielectric constant of the insulating material was considered, and less-complicated structures can be fabricated. At this point, both stacked and trench capacitor structures are dominant in the industry. Indeed, the ability to fabricate deep-trench structures will help guide the need for the incorporation of alternative dielectrics. These structures often require the use of chemical vapor deposition (CVD) methods to ensure conformality over the stacked structure topography or deep within trenches that have large aspect ratios. More recently, atomic-layer CVD processes have been examined for this purpose as well. Although capacitors do not have requirements regarding lateral transport of carriers, as transistors do, capacitors have much more rigorous charge-storage-capacity requirements when they serve as memory elements. This additional constraint results in projected limitations.

28.3.3 Dielectrics for Volatile Memory Capacitors

Capacitor dielectric properties for dynamic random-access memory (DRAM) applications have been important since these devices were introduced commercially in the early 1970s for volatile memory [28.178]. *Fazan* provides some rules of thumb that provide guidance for the selection of desirable materials (and electrical) properties for DRAM capacitor dielectrics [28.176]. In addition to the desire to maintain the storage capacitance around 30 fF/cell, leakage of charge from the storage capacitor must be kept to a cur-

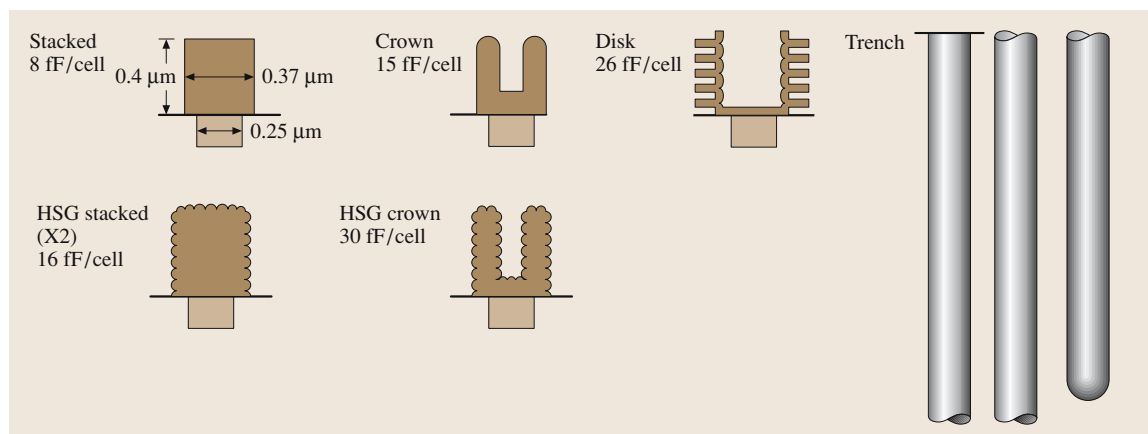


Fig. 28.25 Assorted capacitor structures employed in IC technology. Courtesy of S. Summerfelt

rent density of less than $0.1 \mu\text{A}/\text{cm}^2$ so that less than $\approx 10\%$ of the capacitor charge is lost during the associated refresh (recharge) cycle. Scaling the DRAM cell size has resulted in the consideration of a variety of dielectric materials [28.12, 27, 28].

SiO_2 and SiO_xN_y

The earliest DRAM planar capacitors utilized SiO_2 as the dielectric. Subsequent scaling into the Mbit regime required an increased capacitance density, and thus a higher dielectric constant. So-called oxide/nitride (ON) or oxide/nitride/oxide (ONO) dielectric stacks were incorporated [28.18]. As the names imply, the dielectric consisted of a stack of SiO_2 and Si_3N_4 layers. The incorporation of the Si nitride layer, typically by a deposition method such as chemical vapor deposition, results in an overall increase in the dielectric constant of the stack. Scaling the stacked capacitors has required the development of alternative dielectrics (Al_2O_3 and Ta_2O_5) at the 130-nm node. Nevertheless, according to the 2003 ITRS, Si_3N_4 will be utilized for DRAM capacitor structures to the 45-nm node.

Al_2O_3

Aluminium oxide exhibits a dielectric constant of $\kappa \approx 9$ and a significant band gap of 8.7 eV, as noted in Table 28.4. However, the rate of capacitor scaling and the charge storage per cell appears to require a dielectric constant significantly larger than this value. As a result, substantial research and development was also conducted to incorporate tantalum pentoxide into DRAM capacitor dielectrics. According to the 2003 ITRS, Al_2O_3 will be utilized for DRAM capacitor structures to the 45-nm node.

Ta_2O_5

Amorphous tantalum pentoxide provides a dielectric constant of $\kappa \approx 25$ with a concomitantly smaller band gap (≈ 4.4 eV). Ta_2O_5 films are often deposited by CVD processes for conformality in stacked or trench capacitor structures. Metal–insulator–metal structures are required to preserve the maximum capacitance density, as reactions with polysilicon results in the formation of a thin SiO_2 layer as discussed above in connection with gate dielectric materials. Moreover, the use of MIM structures permits the possibility of a Ta_2O_5 crystalline microstructure, which enables a $\kappa \approx 50$. Chaneleire et al. has summarized the research and development associated with Ta_2O_5 films [28.179]. According to the 2003 ITRS, Ta_2O_5 will also be utilized for DRAM capacitor structures to the 45-nm node.

Barium Strontium Titanate (BST)

Considerable effort has also been exerted in the search for CMOS-compatible “ultra high- κ ” dielectrics. Such materials are envisioned to be required for scaling at and beyond the 45-nm node [28.12]. CVD BST films have been a focus of these efforts in the recent past, with a dielectric constant of $\kappa \approx 250$. The movement of the Ti atom (ion) in the BST lattice structure results in a substantial contribution to the polarization of this materials system. Utilization of this material has required the use of noble-metal electrodes including Pt, Ru (RuO_2) and Ir to control interfacial reactions. As a result, considerable process complexity is introduced into the manufacturing process.

Alternative Dielectric Materials

We also note that recent research on alternative gate dielectric materials, such as HfO_2 and HfSiO , has also rekindled interest in these materials for capacitor applications. The prospect of better interfacial oxide formation control has been one motivating factor, although substantial further work is still required.

28.3.4 Dielectrics for Nonvolatile Memory

As the name implies, nonvolatile memory devices retain their state whether power is applied to the device or not. A thorough description of such devices is provided by Hori [28.18]. For example, the electrically erasable programmable read-only memory (EEPROM) device requires an erase operation prior to programming (*writing*) new data to the device. To accomplish this, a stacked-gate MOS structure is utilized where the intermediate gate is embedded in a dielectric – a so-called *floating gate*. The placement of a higher electric field to permit Fowler–Nordheim tunneling through such dielectrics to the floating gate is utilized to program arrays of these elements – called *flash memory*. Scaling and the reliability required for such nonvolatile memory devices has required the evolution from using SiO_2 to SiO_xN_y .

SiO_2 and SiO_xN_y

For flash memory elements, the formation of nitrided SiO_2 enables a suitable dielectric for reliable operation. The nitridation process, as described earlier, entails the exposure of an SiO_2 dielectric to anneals with N_2 , NH_3 , or N_2O (or combinations of these), often under rapid thermal annealing conditions. Details of this process have been summarized by Hori [28.18]. Subsequent reoxidation of the dielectric is also employed to improve reliability properties.

So-called silicon oxide nitride oxide silicon (SONOS) structures are under examination to improve the scaling limitations of the dielectric thickness and tunneling properties [28.180].

Further scaling of transistors presents significant challenges for flash memory. The thickness of the dielectric is again a key constraint given the need for reliable operation under high-electric-field conditions. Memories based on Si nanocrystals embedded within SiO_2 are currently under development to address this challenge [28.182].

Ferroelectrics

Ferroelectric materials, such as $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$ (PZT) or $\text{SrBi}_2\text{Ta}_2\text{O}_9$ (SBT) have been examined for some time as potential dielectric candidates for memory elements based upon the ferroelectric effect [28.26]. In storage capacitors incorporating such materials, the polarization state of the ferroelectric is preserved (once poled) without the presence of an electric field [28.25]. This state is then sensed by associated circuitry and can therefore be used as a memory device.

Principles. Ferroelectricity describes the spontaneous alignment of dipoles in a dielectric as the result of an externally applied electric field. This alignment behavior has thermal constraints in that heating the dielectric above the Curie temperature results in a phase transformation to a paraelectric state. Therefore, capacitors and circuits which utilize the ferroelectric effect must contain materials that have a relatively high Curie temperature compared to that experienced during operation and reliability testing (generally $> 200^\circ\text{C}$). As shown in Fig. 28.26, the polarization response of these materials to an externally applied electric field exhibits a hys-

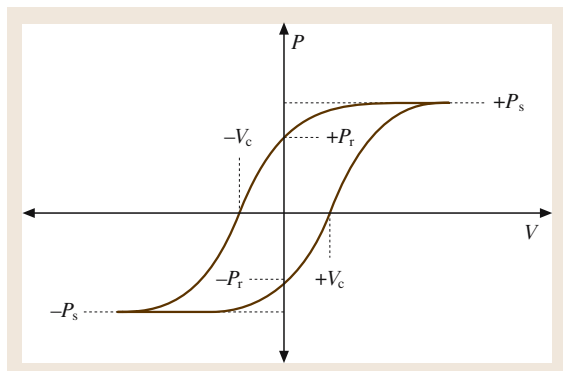


Fig. 28.26 Polarization-voltage curve for a ferroelectric material. After [28.181] (© 2004 IEEE, with permission)

teresis behavior [28.181]. (This behavior is analogous to that observed in ferromagnetic materials, hence the name. Note however that ferroelectric materials contain no iron.) Upon increasing the voltage across the ferroelectric above a coercive value (V_c), the polarization of the material is limited to the spontaneous value (P_s), corresponding to maximum domain alignment. Removal of the electric field results in a decrease of the polarization to the remnant value (P_r), which can be sensed accordingly. The time required for the polarization to be switched is on the order of nanoseconds for these devices. Moreover, this can be accomplished in low-power circuits, making these materials attractive, in principle, for scaled CMOS.

Materials. In addition to the perovskite PZT and SBT materials mentioned above, other materials investigated for nonvolatile memory applications include BaTiO_3 , PbTiO_3 , $\text{Pb}_{1-x}\text{La}_x\text{Zr}_{1-y}\text{Ti}_y\text{O}_3$ (PLZT), $\text{PbMg}_{1-x}\text{Nb}_x\text{O}_3$ (PMN), $\text{SrBi}_2\text{Nb}_2\text{O}_9$ (SBN), and $\text{SrBi}_2(\text{Ta}_{1-x}\text{Nb}_x)_2\text{O}_9$ (SBTN). These materials generally have the cubic structure ABO_3 shown in Fig. 28.27, with the larger (A) cations in the corner of the cubic unit cell, and the smaller (B) cations (e.g. Ti, Zr, Mg, Nb, Ta) located in the center of the unit cell. The oxygen anions are located at the face-centered-cube positions. The distortion of the unit cell in response to the externally applied electric field results in the observed polarization. In particular, the movement of the B cation (e.g. Ti) in the associated lattice between the equilibrium

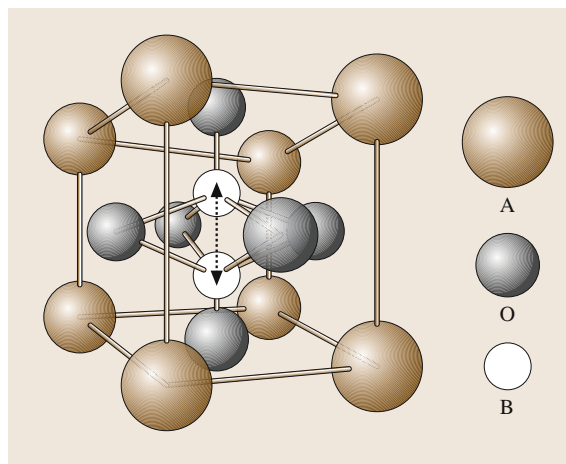


Fig. 28.27 Perovskite ABO_3 unit cell showing motion of the B ion in the cell among the equilibrium positions, resulting in ferroelectric polarization behavior. After [28.181] (© 2004 IEEE, with permission)

positions results in substantial distortion and therefore polarization [28.26, 181].

Issues for Ferroelectric Materials. Substantial research is underway to understand the reliability issues associated with ferroelectric memory devices. The retention of the polarization state is one area of investigation. It is observed that the polarization state decreases slowly over time (*log-time* decay behavior), even in the absence of an electric field, and the reasons are still poorly understood. Another area of concern is *imprint*, where a polarization state, if repeatedly poled to the same state, becomes preferred. Subsequent switching to the opposite state can result in relatively poor retention times. Again, the physical mechanism associated with this phenomenon is poorly understood. The role of hydrogen exposure is another area of investigation, where exposure of ferroelectric random-access memory (FeRAM) elements to hydrogen, commonly from forming gas during CMOS back-end processing, can result in the suppression of

the remnant polarization [28.183–185]. Barriers for hydrogen permeation into the ferroelectric are an area of investigation as well.

Phase-Change Memory

An alternative to charge-storage devices, based upon a controlled phase change, is also now under consideration for scaled integrated circuits [28.186]. Chalcogenides, such as GeSbTe, have been utilized in compact-disk memory storage technology. In that technology, a laser heats a small volume of the material, resulting in a phase change between crystalline and amorphous states, which obviously changes the reflectivity of the exposed region. For the IC application, an electric current is passed through these materials to accomplish the phase-change effect, dramatically altering the resistance of the region. This utilization of alternative (non-dielectric) material is another example of new directions under consideration for IC scaling.

28.4 Interconnect Dielectrics

As noted in Fig. 28.1, the performance of the integrated circuit, as measured by the time delay for signal propagation, also depends upon the interconnections between circuit elements. The scaling of CMOS has resulted in a substantial increase in interconnect metal lines throughout the IC chip, which make a major contribution to the delay time. The industry segments these levels into *local* (interconnection between neighboring devices), *intermediate* (metal 1 interconnection between neighboring circuits), and *global* (interconnection across the chip), as shown in Fig. 28.28. A cross section of a contemporary 65-nm-node IC is shown in Fig. 28.29 where eight layers of metallization and the associated dielectric isolation are evident.

As discussed in Sect. 28.0.1, the *RC* time delay for the interconnect contribution to performance can be attributed to the metal lines and their isolation dielectrics (see Fig. 28.1). The resistivity of the lines has been reduced in the industry by recently adopting copper metallization processes in lieu of aluminium metallization in 1998. Further reductions in the delay time then require the consideration of the dielectric between the lines, as these essentially form a parasitic capacitor structure, and therefore low- κ dielectrics are required. It is noted that for the global interconnect level, new concepts such as RF or optical communication, will likely be needed for continued CMOS scaling.

28.4.1 Tetraethoxysilane (TEOS)

For many years, CVD-deposited SiO₂ provided adequate isolation for interconnection of ICs. This was frequently accomplished through the deposition of tetraethoxysilane (TEOS) and subsequent densification thermal treatments to render a dielectric constant of $\kappa \approx 4$. Films produced in this manner were relatively easy to process and provided good mechanical strength. The incorporation of fluorine into these films [fluorinated silicate glass or (FSG)] succeeded in a reduction of the dielectric constant to $\kappa \approx 3.5$ – 3.7 after considerable efforts. As F is among the most electronegative elements, the incorporation of F into the silica matrix renders the film less polarizable due to Si–F bond formation, and therefore results in a lower permittivity. Scaling CMOS however has driven the industry to consider interconnect dielectric materials with $\kappa \ll 4$.

28.4.2 Low- κ Dielectrics

As seen in Figures 28.28 and 28.29, several levels of dielectrics must be incorporated with the metallization schemes associated with ICs. As can be seen, dielectrics utilized in this back end of (fabrication) line (BEOL) portion of the IC fabrication process are also segmented into pre-, inter- and intra-metal dielectrics. Addition-

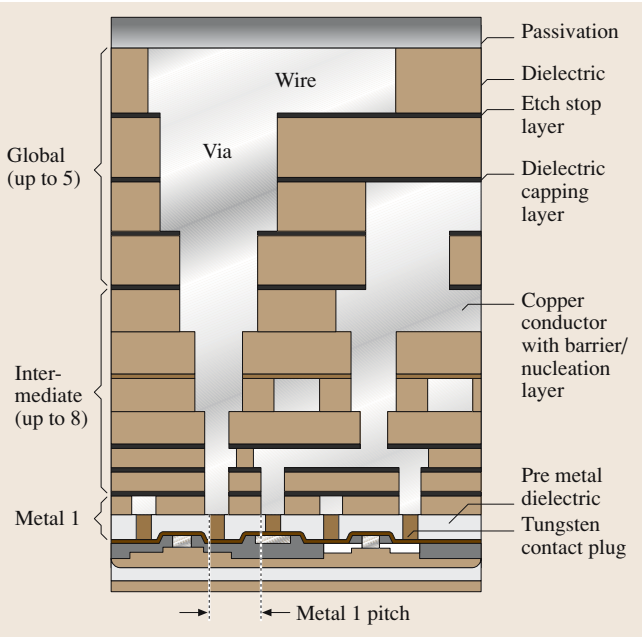


Fig. 28.28 Schematic of a typical chip cross section showing key interconnect metallization and dielectric layers. From [28.12]

ally, dielectric layers are employed to facilitate etching control for Cu metal patterning of the various interconnection lines. This requires materials integration with barrier layers (that control Cu diffusion), etch stop layers, CMP (chemical mechanical polishing) stop layers as well as overall mechanical stability. Additionally, for the local and metal 1 layers, the future incorporation of metal gate and/or NiSi materials may require extensive limitations on processing temperatures ($< 500^{\circ}\text{C}$).

Several materials are under investigation for the low- κ application, and some of these are summarized in Table 28.5 [28.12, 28]. Most consist of polymeric (low-Z) materials with varying porosity

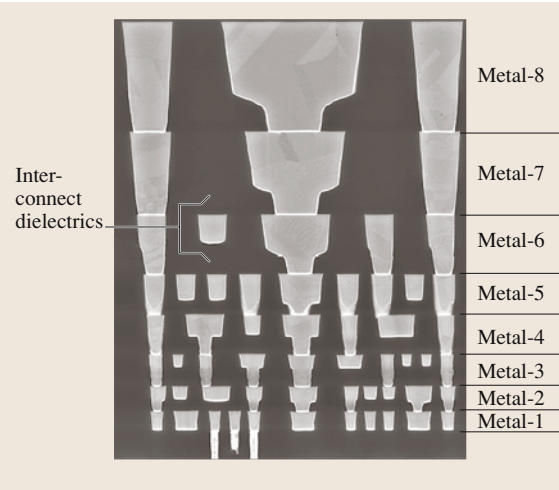


Fig. 28.29 Cross section of a modern IC chip for the 65-nm node to be in production in 2005 showing eight levels of metal interconnects isolated by low- κ dielectrics. (Courtesy of Intel)

to enable a sufficiently low dielectric constant for interconnect applications. Recent technology announcements for the 90-nm and 65-nm nodes often refer to these materials as carbon-doped oxides [28.187, 188]. At this point, films are deposited by either CVD or spin-on methods. Desirable properties for low- κ dielectrics include thermal stability (to $\approx 500^{\circ}\text{C}$), mechanical stability (to withstand packaging), electrical isolation stability/reliability (similar to SiO_2), chemical stability (minimal moisture absorption, resistance to process chemicals, resist compatibility), compatibility with BEOL materials (e.g., diffusion barriers) and processes (etching, cleaning, post-metallization and forming-gas anneals, chemical mechanical polishing, etc.) and of course low cost. As noted by Wolf, thermal properties of low- κ materials are important

Table 28.5 Interconnect dielectric materials (after [28.12, 28])

Dielectric constant	Material	Pre-production year
3.5–3.7	FSG	Current
3.0–3.6	Polyimides	Current
2.7–3.1	Spin-on glass	Current
2.6–2.9	Organo-silicate glasses (OSG)	Current – 2011
2.6–2.8	Parylene-based polymers	Current
2.5–3.2	Methyl/hydrogen silsesquioxane (MSQ/HSQ)	2005
1.8–2.4	Porous MSQ, parylene-based polymers	2008
1.1–2.2	Silica aerogels	2011
1.5–2.2	Silica xerogels	2011

for effective power dissipation in high-performance applications.

According to the ITRS roadmap, deposited silicon oxides are envisioned to be employed for pre-metal dielectrics through the 45-nm node (≈ 2012). The introduction of Ni-silicide contacts, metal gate electrodes and high- κ dielectrics will certainly have an impact on the development of alternatives to these deposited oxides. Methyl/hydrogen silsesquioxane (MSQ/HSQ) appear to be under consideration for further development in this regard.

28.5 Summary

It should be evident that IC technology is critically dependent upon suitable dielectrics throughout the entire chip. Materials properties, and their resultant electrical properties, must be carefully evaluated throughout the research and development process associated with integrated circuit technology.

For inter/intra-metal dielectrics, a wide array of materials are envisioned, as seen in Table 28.5. Both FSG and organo-silicate glasses (OSG) are envisioned to address the requirements of these interconnect dielectrics to the 45-nm node. Thereafter, alternative materials ($\kappa < 2.8$) mentioned in Table 28.5 will need to be developed.

Patterning (hard mask) and etch-stop dielectrics will continue to employ Si-oxides, Si-nitrides, Si-oxycarbides, and Si-carbonitrides to the 45-nm node. Thereafter, alternative materials for patterning may be required.

Recent years and roadmap predictions clearly place an emphasis on the development of new materials for the various dielectrics employed to achieve scaling. Researchers and technologists engaged in this endeavor must be able to span several disciplines to enable the successful integration of these new dielectric materials.

References

- 28.1 G. Wilk, R. M. Wallace, J. M. Anthony: J. Appl. Phys. **89**(10), 5243 (2001)
- 28.2 R. M. Wallace, G. Wilk: Critical Rev. Solid State Mater. Sci. **28**, 231 (2003)
- 28.3 R. M. Wallace: Appl. Surf. Sci. **231–232**, 543 (2004)
- 28.4 H.-S. P. Wong: *ULSI Devices*, ed. by C. Y. Chang, S. M. Sze (Wiley, New York 2000) Chap. 3
- 28.5 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 3 (Lattice, Sunset Beach 1995)
- 28.6 E. Nicollian, J. Brews: *MOS Physics and Technology* (Wiley, New York 1982)
- 28.7 S. M. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)
- 28.8 G. Moore: Electronics **38**, 8 (1965). Also see: <http://www.intel.com/labs/index.htm>
- 28.9 G. Moore: *Tech. Dig. Int. Electron. Dev. Meet* (IEEE, Washington, D.C. 1975) p. 11
- 28.10 P. E. Ross: IEEE Spectrum **40**(12), 30 (2003)
- 28.11 Intel: *Expanding Moore's Law*; see: <http://www.intel.com/labs/index.htm> (2002)
- 28.12 ITRS, see: <http://public.itrs.net/> (2003)
- 28.13 H. Iwai: Microelec. Eng. **48**, 7 (1999)
- 28.14 H. Iwai, H. S. Monose, S.-I. Ohmi: *The Physics and Chemistry of SiO₂ and the Si–SiO₂ Interface*, Vol. 4, ed. by H. Z. Massoud, I. J. R. Baumvol, M. Hirose, E. H. Poindexter (The Electrochemical Society, Pennington 2000) p. 1
- 28.15 G. Moore: *No Exponential is Forever ... But We Can Delay "Forever"*, keynote address at Int. Solids State Circuits Conference. See: <http://www.intel.com/labs/eml/doc.htm> (2003)
- 28.16 S. C. Sun: IEEE Tech. Dig. Int. Electron. Dev. Meet. Washington, DC, 765 (1997)
- 28.17 M. T. Bohr: IEEE Tech. Dig. Int. Electron. Dev. Meet. Washington, DC, 241 (1995)
- 28.18 T. Hori: *Gate Dielectrics and MOS ULSI's*, Series in Electronics and Photonics, Vol. 34 (Springer, Berlin 1997)
- 28.19 A. Chatterjee, M. Rodder, I.-C. Chen: IEEE Trans. Electron. Dev. **45**, 1246 (1998)
- 28.20 I.-C. Chen, W. Liu: *ULSI Devices*, ed. by C. Y. Chang, S. M. Sze (Wiley, New York 2000) Chap. 10
- 28.21 R. Dennard, F. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, A. R. LeBlanc: J. Solid State Circuits **SC-9**, 256 (1974)
- 28.22 R. Dennard, F. Gaensslen, E. Walker, P. Cook: J. Solid State Circuits **SC-14**, 247 (1979)
- 28.23 D. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, H.-S. P. Wong: Proc. IEEE **89**, 259 (2001)
- 28.24 S. Thompson, P. Packan, M. Bohr: Intel Technol. J. **Q 3**, 223–225 (1998)
- 28.25 S. O. Kasap: *Principles of Electrical Engineering Materials and Devices* (McGraw-Hill, New York 2002)
- 28.26 R. Ramesh: *Thin Film Ferroelectric Materials and Devices* (Kluwer, Boston 1997)

- 28.27 D.-S. Yoon, J. S. Roh, H. K. Baik, S.-M. Lee: *Crit. Rev. Solid State Mater. Sci.* **27**, 143 (2002)
- 28.28 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 4 (Lattice, Sunset Beach 2002)
- 28.29 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 1 (Lattice, Sunset Beach 1986)
- 28.30 S. Wolf: *Silicon Processing for the VLSI Era*, Vol. 2 (Lattice, Sunset Beach 1990)
- 28.31 B. Bivari: IEEE Tech. Dig. Int. Electron. Dev. Meet., 555 (1996)
- 28.32 S. Banerjee, B. Streetman: *ULSI Devices*, ed. by C. Y. Chang, S. M. Sze (Wiley, New York 2000) Chap. 4
- 28.33 R. Rios, N. D. Arora: IEEE Tech. Dig. Int. Electron. Dev. Meet. San Francisco, 613 (1994)
- 28.34 A. C. Diebold, D. Venables, Y. Chabal, D. Muller, M. Weldon, E. Garfunkel: *Mater. Sci. Semicond. Proc.* **2**, 104 (1999)
- 28.35 Z. H. Lu, J. P. McCaffrey, B. Brar, G. D. Wilk, R. M. Wallace, L. C. Feldman, S. P. Tay: *Appl. Phys. Lett.* **71**, 2764 (1997)
- 28.36 Y.-C. King, C. Hu, H. Fujioka, S. Kamohara: *Appl. Phys. Lett.* **72**, 3476 (1998)
- 28.37 K. Yang, Y.-C. King, C. Hu: *Symp. VLSI Tech. Tech. Dig. Papers*, Kyoto, Japan, 77 (1999)
- 28.38 C. Y. Wong, J. Y. Sun, Y. Taur, C. S. Oh, R. Angelucci, B. Davari: IEEE Tech. Dig. Int. Electron. Dev. Meet. San Francisco, 238 (1988)
- 28.39 E. M. Vogel, K. Z. Ahmed, B. Hornung, W. K. Henson, P. K. McLarty, G. Lucovsky, J. R. Hauser, J. J. Wortman: IEEE Trans. Electron. Dev. **45**, 1350 (1998)
- 28.40 M. L. Green, E. P. Gusev, R. Degraeve, E. L. Garfunkel: *J. Appl. Phys.* **90**, 2057 (2001)
- 28.41 D. K. Schroder: *Semiconductor Material and Device Characterization*, 2nd edn. (Wiley, New York 1998)
- 28.42 R. Hummel: *Electronic Properties of Materials*, 2nd edn. (Springer, New York 1993)
- 28.43 W. R. Runyan, K. E. Bean: *Semiconductor Integrated Circuit Processing Technology* (Addison-Wesley, New York 1990)
- 28.44 S. T. Pantilides: *The Physics of SiO₂ and its Interfaces* (Pergamon, New York 1978)
- 28.45 G. Lucovsky, S. T. Pantilides, F. L. Galeener: *The Physics of MOS Insulators* (Pergamon, New York 1980)
- 28.46 C. R. Helms, B. E. Deal: *The Physics and Chemistry of SiO₂ and the Si-SiO₂ Interface* (Plenum, New York 1988)
- 28.47 P. Balk (ed): *The Si-SiO₂ System*, *Mater. Sci. Monogr.* (Elsevier, New York 1988) p. 32
- 28.48 C. R. Helms, B. E. Deal: *The Physics and Chemistry of SiO₂ and the Si-SiO₂ Interface*, 2 (Plenum, New York 1993)
- 28.49 H. Z. Massoud, E. H. Poindexter, C. R. Helms: *The Physics of SiO₂ and its Interfaces - 3*, Vol. 96-1 (Electrochemical Society, Pennington 1996)
- 28.50 R. A. B. Devine: *The Physics and Technology of Amorphous SiO₂* (Plenum, New York 1988) p. 2
- 28.51 E. Garfunkel, E. Gusev, A. Vul': *Fundamental Aspects of Ultrathin Dielectrics on Si-based Devices*, NATO Science Series, Vol. 3147 (Kluwer, Nowell 1998)
- 28.52 R. A. B. Devine, J.-P. Duraud, E. Dooryhee: *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide* (Wiley, New York 2000)
- 28.53 H. Z. Massoud, I. J. R. Baumvol, M. Hirose, E. H. Poindexter: *The Physics of SiO₂ and its Interfaces - 4*, Vol. PV2000-2 (Electrochemical Society, Pennington 2000)
- 28.54 C. R. Helms, E. H. Poindexter: *Rep. Prog. Phys.* **57**, 791 (1994)
- 28.55 G. Dolino: *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide*, ed. by R. A. B. Devine, J.-P. Duraud, E. Dooryhee (Wiley, New York 2000) Chap. 2
- 28.56 L. W. Hobbs, C. E. Jesurum, B. Berger: *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide*, ed. by R. A. B. Devine, J.-P. Duraud, E. Dooryhee (Wiley, New York 2000) Chap. 1
- 28.57 A. Bongiorno, A. Pasquarello: *Appl. Phys. Lett.* **83**, 1417 (2003)
- 28.58 F. Mauri, A. Pasquarello, B. G. Pfrommer, Y.-G. Yoon, S. G. Louie: *Phys. Rev. B* **62**, R4786 (2000)
- 28.59 W. H. Zachariasen: *J. Am. Chem. Soc.* **54**, 3841 (1932)
- 28.60 P. Balk: *J. Nanocryst. Sol.* **187**, 1-9 (1995)
- 28.61 F. J. Grunthaner, P. J. Grunthaner: *Mater. Sci. Rep.* **1**, 65 (1986)
- 28.62 A. Bongiorno, A. Pasquarello: *Mater. Sci. Eng. B* **96**, 102 (2002)
- 28.63 A. Stirling, A. Pasquarello: *Phys. Rev. B* **66**, 24521 (2002)
- 28.64 G. Hollinger, F. R. Himpel: *Phys. Rev. B* **28**, 3651 (1983)
- 28.65 G. Hollinger, F. R. Himpel: *Appl. Phys. Lett.* **44**, 93 (1984)
- 28.66 F. J. Himpel, F. R. McFeely, A. Taleb-Ibrahimi, J. A. Yarmoff, G. Hollinger: *Phys. Rev. B* **38**, 6084 (1988)
- 28.67 T. Hattori: *Crit. Rev. Solid State Mater. Sci.* **20**, 339 (1995)
- 28.68 F. Rochet, Ch. Poncey, G. Dufour, H. Roulet, C. Guillot, F. Sirotti: *J. Non-Cryst. Solids* **216**, 148 (1997)
- 28.69 J. H. Oh, H. W. Yeom, Y. Hagimoto, K. Ono, M. Oshima, N. Hirashita, M. Nywa, A. Toriumi, A. Kakizaki: *Phys. Rev. B* **63**, 205310 (2001)
- 28.70 D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, G. Timp: *Nature* **399**, 758 (1999)
- 28.71 I. J. R. Baumvol: *Surf. Sci. Rep.* **36**, 1 (1999)
- 28.72 J. J. Lander, J. Morrison: *J. Appl. Phys.* **33**, 2089 (1962)
- 28.73 F. W. Smith, G. Ghidini: *J. Electrochem. Soc.* **129**, 1300 (1982)
- 28.74 K. Wurm, R. Kliese, Y. Hong, B. Röttger, Y. Wei, H. Neddermeyer, I. S. T. Tsong: *Phys. Rev. B* **50**, 1567 (1994)
- 28.75 J. Seiple, J. P. Pelz: *Phys. Rev. Lett.* **73**, 999 (1994)

- 28.76 J. Seiple, J. P. Pelz: J. Vac. Sci. Technol. A **13**, 772 (1995)
- 28.77 Y. Wei, R. M. Wallace, A. C. Seabaugh: J. Appl. Phys. **81**, 6415 (1997)
- 28.78 E. P. Gusev, H. C. Lu, T. Gustafsson, E. Garfunkel: Phys. Rev. B **52**, 1759 (1995)
- 28.79 B. E. Deal, A. S. Grove: J. Appl. Phys. **36**, 3770 (1965)
- 28.80 J. D. Plummer: Silicon oxidation kinetics—from Deal–Grove to VLSI process models. In: *The Physics of SiO₂ and its Interfaces* – 3, Vol. 96–1, ed. by H. Z. Massoud, E. H. Poindexter, C. R. Helms (Electrochemical Society, Pennington 1996) p. 129
- 28.81 P. Balk: Trans. IEEE **53**, 2133 (1965)
- 28.82 G. Abowitz, E. Arnold, J. Ladell: Phys. Rev. Lett. **18**, 543 (1967)
- 28.83 B. E. Deal, M. Sklar, A. S. Grove, E. H. Snow: J. Electrochem. Soc. **114**, 266 (1967)
- 28.84 E. Arnold, J. Ladell, G. Abowitz: Appl. Phys. Lett. **13**, 413 (1968)
- 28.85 R. R. Razouk, B. E. Deal: J. Electrochem. Soc. **126**(9), 1573–1581 (Sept. 1979)
- 28.86 P. J. Caplan, E. H. Poindexter, B. E. Deal, R. R. Razouk: J. Appl. Phys. **50**, 5847 (1979)
- 28.87 P. J. Caplan, E. H. Poindexter, B. E. Deal, R. R. Razouk: *The Physics of MOS Insulators*, ed. by G. Lucovsky, S. T. Pantilides, F. L. Galeener (Pergamon, New York 1980) p. 306
- 28.88 J. H. Weil, J. R. Bolton, J. E. Wertz: *Electron Paramagnetic Resonance: Elementary Theory and Practical Applications* (Wiley, New York 1994)
- 28.89 P. J. Caplan, J. N. Helbert, B. E. Wagner, E. H. Poindexter: Surf. Sci. **54**, 33 (1976)
- 28.90 E. H. Poindexter, P. J. Caplan: Prog. Surf. Sci. **14**, 201 (1983)
- 28.91 E. H. Poindexter, P. J. Caplan: J. Vac. Sci. Technol. A **6**, 390 (1988)
- 28.92 J. F. Conley: Mater. Res. Soc. Symp. Proc. **428**, 293 (1996)
- 28.93 J. F. Conley, P. M. Lenahan: A review of electron spin resonance spectroscopy of defects in thin film SiO₂ on Si. In: *The Physics of SiO₂ and its Interfaces* – 3, Vol. 96–1, ed. by H. Z. Massoud, E. H. Poindexter, C. R. Helms (Electrochemical Society, Pennington 1996) p. 214
- 28.94 P. M. Lenahan, J. F. Conley: J. Vac. Sci. Technol. B **16**, 2134 (1998)
- 28.95 A. G. Revesz, B. Goldstein: Surf. Sci. **14**, 361 (1969)
- 28.96 Y. Nishi: J. Appl. Phys. **10**, 52 (1971)
- 28.97 I. Shiota, N. Miyamoto, J.-I. Nishizawa: J. Appl. Phys. **48**, 2556 (1977)
- 28.98 E. H. Poindexter, E. R. Ahlstrom, P. J. Caplan: *The Physics of SiO₂ and its Interfaces*, ed. by S. T. Pantilides (Pergamon, New York 1978) p. 227
- 28.99 G. J. Gerardi, E. H. Poindexter, P. J. Caplan, N. M. Johnson: Appl. Phys. Lett. **49**, 348 (1986)
- 28.100 D. Sands, K. M. Brunson, M. H. Tayanani–Najaran: Semicond. Sci. Technol. **7**, 1091 (1992)
- 28.101 K. L. Brower: Phys. Rev. B **38**, 9657 (1988)
- 28.102 K. L. Brower: Phys. Rev. B **42**, 3444 (1990)
- 28.103 A. Stesmans, V. V. Afanas'ev: J. Appl. Phys. **83**, 2449 (1998)
- 28.104 A. Stesmans, V. V. Afanas'ev: J. Phys. Condens. Matter **10**, L19 (1998)
- 28.105 A. Stesmans, V. V. Afanas'ev: Micro. Eng. **48**, 116 (1999)
- 28.106 A. Stesmans, B. Nouwen, V. Afanas'ev: Phys. Rev. B **58**, 15801 (1998)
- 28.107 E. H. Poindexter, P. J. Caplan: *Insulating Films on Semiconductors*, ed. by M. Schulz, G. Pensl (Springer, Berlin, Heidelberg 1981) p. 150
- 28.108 M. Schulz, G. Pensl: *Insulating Films on Semiconductors* (Springer, New York 1981)
- 28.109 W. H. Brattain, J. Bardeen: Bell Syst. Tech. J. **13**, 1 (1953)
- 28.110 W. L. Brown: Phys. Rev. **91**, 518–527 (1953)
- 28.111 E. N. Clarke: Phys. Rev. **91**, 756 (1953)
- 28.112 H. R. Huff: J. Electrochem. Soc. **149**, S35 (2002)
- 28.113 J. T. Law: J. Phys. Chem. **59**, 67 (1955)
- 28.114 J. T. Law, E. E. Francois: J. Phys. Chem. **60**, 353 (1956)
- 28.115 J. T. Law: J. Phys. Chem. **61**, 1200 (1957)
- 28.116 J. T. Law: J. Appl. Phys. **32**, 600 (1961)
- 28.117 P. Balk: Microelectron. Eng. **48**, 3 (1999)
- 28.118 E. Kooi: Philips Res. Rep. **20**, 578 (1965)
- 28.119 P. Balk: Electrochem. Soc. Ext. Abstracts **14**(109), 237 (1965)
- 28.120 P. Balk: Electrochem. Soc. Ext. Abs. **14**(111), 29 (1965)
- 28.121 P. Balk: J. Electrochem. Soc. **112**, 185C (1965d)
- 28.122 E. Kooi: Philips Res. Rep. **21**, 477 (1966)
- 28.123 A. Stesmans: Appl. Phys. Lett. **68**, 2076 (1996)
- 28.124 A. Stesmans, V. Afanas'ev: Micro. Eng. **36**, 201 (1997)
- 28.125 K. L. Brower, S. M. Myers: Appl. Phys. Lett. **57**, 162 (1999)
- 28.126 A. Stesmans: Phys. Rev. B **48**, 2418 (1993)
- 28.127 A. G. Revesz: J. Electrochem. Soc. **126**, 122 (1979)
- 28.128 G. J. Gerardi, E. H. Poindexter: J. Electrochem. Soc. **136**, 588 (1989)
- 28.129 T. R. Oldham, F. B. McLean, H. E. Jr. Boesch, J. M. McGarrity: Semicond. Sci. Technol. **4**, 986 (1989)
- 28.130 A. Stesmans, V. Afanas'ev: Phys. Rev. B **54**, 11129 (1996)
- 28.131 E. Wu, B. Linder, J. Stathis, W. Lai: IEEE Tech. Dig. Int. Electron. Dev. Meet. Washington, DC, 919 (2003)
- 28.132 D. A. Buchanan: IBM J. Res. Devel. **43**, 245 (1999)
- 28.133 Y. Wu, G. Lucovsky, Y.–M. Lee: IEEE Trans. Electron. Dev. **47**, 1361 (2000)
- 28.134 M. Hillert, S. Jonsson, B. Sundman: Z. Metallkd. **83**, 648 (1992)
- 28.135 D. M. Brown, P. V. Gray, F. K. Heumann, H. R. Philipp, E. A. Taft: J. Electrochem. Soc. **115**, 311 (1968)
- 28.136 E. H. Poindexter, W. L. Warren: J. Electrochem. Soc. **142**, 2508 (1995)
- 28.137 J. T. Yount, P. M. Lenahan, P. W. Wyatt: J. Appl. Phys. **74**, 5867 (1993)

- 28.138 K. Kushida–Abdelghafar, K. Watanabe, T. Kikawa, Y. Kamigaki, J. Ushio: *J. Appl. Phys.* **92**, 2475 (2002)
- 28.139 G. Lucovsky, T. Yasuda, Y. Ma, S. Hattangady, V. Misra, X.–L. Xu, B. Hornung, J.J. Wortman: *J. Non–Cryst. Solids* **179**, 354 (1994)
- 28.140 S. V. Hattangady, H. Niimi, G. Lucovsky: *Appl. Phys. Lett.* **66**, 3495 (1995)
- 28.141 S. V. Hattangady, R. Kraft, D. T. Grider, M. A. Douglas, G. A. Brown, P. A. Tiner, J. W. Kuehne, P. E. Nicollian, M. F. Pas: *IEEE Tech. Dig. Int. Electron. Dev. Meet. San Francisco*, 495 (1996)
- 28.142 H. Yang, G. Lucovsky: *Tech. Dig. Int. Electron. Dev. Meet. Washington, DC*, 245 (1999)
- 28.143 J. P. Chang, M. L. Green, V. M. Donnelly, R. L. Opila, J. Eng Jr., J. Sapjeta, P. J. Silverman, B. Weir, H. C. Lu, T. Gustafsson, E. Garfunkel: *J. Appl. Phys.* **87**, 4449 (2000)
- 28.144 R. M. Wallace, G. Wilk: *Mater. Res. Soc. Bull.*, 192 (March 2002) also see this focus issue for reviews of other aspects on gate dielectric issues
- 28.145 H. Iwai, S. Ohmi, S. Akama, C. Ohshima, A. Kikuchi, I. Kashiwagi, J. Taguchi, H. Yamamoto, J. Tonotani, Y. Kim, I. Ueda, A. Kuriyama, Y. Yoshihara: *IEEE Tech. Dig. Int. Electron. Dev. Meet. San Francisco*, 625 (2002)
- 28.146 D. G. Schlom, J. H. Haeni: *Mater. Res. Soc. Bull.* **27**(3), 198 (2002) and refs. therein
- 28.147 Y.–C. Yeo: *Thin Solids Films* **462–3**, 34 (2004) and references therein
- 28.148 S. Guha, E. Cartier, N. A. Bojarczuk, J. Bruley, L. Gignac, J. Karasinski: *J. Appl. Phys.* **90**, 512 (2001)
- 28.149 S. Miyazaki: *J. Vac. Sci. Technol. B* **19**, 2212 (2001)
- 28.150 J. Robertson, C. W. Chen: *Appl. Phys. Lett.* **74**, 1168 (1999)
- 28.151 J. Robertson: *J. Vac. Sci. Technol. B* **18**, 1785 (2000)
- 28.152 J. Robertson: *J. Non–Cryst. Solids* **303**, 94 (2002)
- 28.153 H. Y. Yu, M. F. Li, B. J. Cho, C. C. Yeo, M. S. Joo, D.–L. Kwong, J. S. Pan, C. H. Ang, J. Z. Zheng, S. Ramanathan: *Appl. Phys. Lett.* **81**, 376 (2002)
- 28.154 E. Zhu, T. P. Ma, T. Tamagawa, Y. Di, J. Kim, R. Caruthers, M. Gibson, T. Furukawa: *IEEE Tech. Dig. Int. Electron. Dev. Meet. (IEEE, Washington, D.C. 2001)* p. 20.4.1.
- 28.155 G. D. Wilk, M. L. Green, M.–Y. Ho, B. W. Busch, T. W. Sorsch, F. P. Klemens, B. Brijs, R. B. van Dover, A. Kornblit, T. Gustafsson, E. Garfunkel, S. Hilleenius, D. Monroe, P. Kalavade, J. M. Hergenrother: *IEEE Tech. Dig. VLSI Symp. Honolulu*, 88 (2002)
- 28.156 H. Nohira, W. Tsai, W. Besling, E. Young, J. Petry, T. Conard, W. Vandervorst, S. De Gendt, M. Heyns, J. Maes, M. Tuominen: *J. Non–Cryst. Solids* **303**, 83 (2002)
- 28.157 M. R. Visokay, J. J. Chambers, A. L. P. Rotondaro, A. Shanware, L. Colombo: *Appl. Phys. Lett.* **80**, 3183 (2002)
- 28.158 M. S. Akbar, S. Gopalan, H.–J. Cho, K. Onishi, R. Choi, R. Nieh, C. S. Kang, Y. H. Kim, J. Han, S. Krishnan, J. C. Lee: *Appl. Phys. Lett.* **82**, 1757 (2003)
- 28.159 K. Torii, T. Aoyama, S. Kamiyama, Y. Tamura, S. Miyazaki, H. Kitajima, T. Arikado: *Tech. Dig. VLSI Symp. Honolulu*, 112 (2004)
- 28.160 M. V. Fischetti, D. A. Nuemayer, E. A. Cartier: *J. Appl. Phys.* **90**, 4587 (2001)
- 28.161 Z. Ren, M. V. Fischetti, E. P. Gusev, E. A. Cartier, M. Chudzik: *IEEE Tech. Dig. Int. Electron. Dev. Meet. Washington, DC*, 793 (2003)
- 28.162 R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, M. Metz: *IEEE Electron. Dev. Lett.* **25**, 408 (2004)
- 28.163 G. D. Wilk, R. M. Wallace: *Appl. Phys. Lett.* **74**, 2854 (1999)
- 28.164 M. Quevedo–Lopez, M. El–Bouanani, S. Addepalli, J. L. Duggan, B. E. Gnade, M. R. Visokay, M. J. Bevan, L. Colombo, R. M. Wallace: *Appl. Phys. Lett.* **79**, 2958 (2001)
- 28.165 S. Guha, E. P. Gusev, H. Okorn–Schmidt, M. Copel, L. Å. Ragnarsson, N. A. Bojarczuk: *Appl. Phys. Lett.* **81**, 2956 (2002)
- 28.166 M. Quevedo–Lopez, M. El–Bouanani, S. Addepalli, J. L. Duggan, B. E. Gnade, M. R. Visokay, M. Douglas, L. Colombo, R. M. Wallace: *Appl. Phys. Lett.* **79**, 4192 (2001)
- 28.167 R. M. C. de Almeida, I. J. R. Baumvol: *Surf. Sci. Rep.* **49**, 1 (2003)
- 28.168 R. M. Wallace, R. A. Stolz, G. D. Wilk: *Zirconium and/or hafnium silicon–oxynitride gate*, US Patent 6 013 553; 6 020 243; 6 291 866; 6 291 867 (2000)
- 28.169 G. D. Wilk, R. M. Wallace, J. M. Anthony: *J. Appl. Phys.* **87**, 484 (2000)
- 28.170 A. L. P. Rotondaro, M. R. Visokay, J. J. Chambers, A. Shanware, R. Khamankar, H. Bu, R. T. Laaksonen, L. Tsung, M. Douglas, R. Kuan, M. J. Bevan, T. Grider, J. McPherson, L. Colombo: *Symp. VLSI Technol. Tech. Dig. Papers, Honolulu*, 148 (2002)
- 28.171 M. Quevedo–Lopez, M. El–Bouanani, M. J. Kim, B. E. Gnade, M. R. Visokay, A. LiFatou, M. J. Bevan, L. Colombo, R. M. Wallace: *Appl. Phys. Lett.* **81**, 1609 (2002)
- 28.172 M. Quevedo–Lopez, M. El–Bouanani, M. J. Kim, B. E. Gnade, M. R. Visokay, A. LiFatou, M. J. Bevan, L. Colombo, R. M. Wallace: *Appl. Phys. Lett.* **82**, 4669 (2003)
- 28.173 M.–Y. Ho, H. Gong, G. D. Wilk, B. W. Busch, M. L. Green, W. H. Lin, A. See, S. K. Lahiri, M. E. Loomans, P. I. Räisänen, T. Gustafsson: *Appl. Phys. Lett.* **81**, 4218 (2002)
- 28.174 Y.–C. Yeo, T.–J. King, C. Hu: *J. Appl. Phys.* **92**, 7266 (2002)
- 28.175 I. S. Jeon, J. Lee, P. Zhao, P. Sivasubramani, T. Oh, H. J. Kim, D. Cha, J. Huang, M. J. Kim, B. E. Gnade, J. Kim, R. M. Wallace: *IEEE Tech. Dig. Int. Electron. Dev. Meet. (IEEE, San Francisco 2004)*
- 28.176 P. C. Fazan: *Integr. Ferroelectr.* **4**, 247 (1994)

- 28.177 H. Schichijo: *ULSI Devices*, ed. by C.Y. Chang, S. M. Sze (Wiley, New York 2000) Chap. 7
- 28.178 P. J. Harrop, D. S. Campbell: *Thin Solid Films* **2**, 273 (1968)
- 28.179 C. Chaneliere, J. L. Autran, R. A. B. Devine, B. Bal-land: *Mater. Sci. Eng. R* **22**, 269 (1998)
- 28.180 S. S. Chung, P.-Y. Chiang, G. Chou, C.-T. Huang, P. Chen, C.-H. Chu, C. C.-H. Hsu: *Tech. Dig. Int. Electron. Dev. Meet.* (IEEE, Washington, D.C. 2003) p. 26.6.1.
- 28.181 G. F. Derbenwick, A. F. Isaacson: *IEEE Circuits Dev.*, 20 (2001)
- 28.182 B. De Salvo: *Tech. Dig. Int. Electron. Dev. Meet.* (IEEE, Washington, D.C. 2003) p. 26.1.1.
- 28.183 A. R. Krauss, A. Dhote, O. Auciello, J. Im, R. Ramesh, A. Aggarwal: *Integr. Ferroelectr.* **27**, 147 (1999)
- 28.184 J. Im, O. Auciello, A. R. Krauss, D. M. Gruen, R. P. H. Chang, S. H. Kim, A. I. Kingon: *Appl. Phys. Lett.* **74**, 1162 (1999)
- 28.185 N. Poonawala, V. P. Dravid, O. Auciello, J. Im, A. R. Krauss: *J. Appl. Phys.* **87**, 2227 (2000)
- 28.186 S. Lai: *Tech. Dig. Int. Electron. Dev. Meet.* (IEEE, Washington, D.C. 2003) p. 10.1.1
- 28.187 S. Thompson: *IEEE Tech. Dig. Int. Electron. Dev. Meet.* (IEEE, San Francisco 2002) p. 765
- 28.188 M. Bohr: 65 nm Press Release, August 2004, www.intel.com/research/silicon

Thin Films

29. Thin Films

This chapter provides an extended introduction to the basic principles of thin-film technology, including deposition processes, structure, and some optical and electrical properties relevant to this volume. The material is accessible to scientists and engineers with no previous experience in this field, and contains extensive references to both the primary literature and earlier review articles. Although it is impossible to provide full coverage of all areas or of the most recent developments in this survey, references are included to enable the reader to access the information elsewhere, while the coverage of fundamentals will allow this to be appreciated.

Deposition of thin films by the main physical deposition methods of vacuum evaporation, molecular-beam epitaxy and sputtering are described in some detail, as are those by the chemical deposition methods of electrodeposition, chemical vapour deposition and the Langmuir–Blodgett technique. Examples of structural features of some thin films are given, including their crystallography, larger-scale structure and film morphology. The dependence of these features on the deposition conditions are stressed, including those required for the growth of epitaxial films and the use of zone models in the classification of the morphological characteristics. The main optical properties of thin films are reviewed, including the use of Fresnel coefficients at

29.1 Deposition Methods	661
29.1.1 Physical Deposition Methods	661
29.1.2 Chemical Deposition Methods	677
29.2 Structure	682
29.2.1 Crystallography	682
29.2.2 Film Structure	683
29.2.3 Morphology	688
29.3 Properties	692
29.3.1 Optical Properties	692
29.3.2 Electrical Properties	696
29.4 Concluding Remarks	708
References	711

media boundaries, reflectance and transmittance, matrix methods and the application of these techniques to the design of antireflection coatings, mirrors and filters. The dependence of electrical conductivity (or resistivity) and the temperature coefficient of resistivity in metallic thin films is discussed, in particular the models of Thomson, Fuchs–Sondheimer and the grain-boundary model of Mayadas–Shatzkes. For insulating and semiconducting thin films the origin and effects of several high-field conduction processes are examined, including space-charge-limited conductivity, the Poole–Frenkel effect, hopping, tunnelling and the Schottky effect. Finally, some speculations regarding future developments are made.

The earliest use of thin films by mankind is probably in the application of glazes to brickwork and pottery. Tin glazing was discovered by the Assyrians and was used to coat decorative brickwork. Glazed brick panels have been recovered from various archeological sites, with a fragment from Nimrud being dated to about 890 B.C. Various alternative types of glaze have since been developed, and are used both to overcome the effects of porosity in pottery and for decorative purposes. The importance of thin films in optics was first recognised in the 17th century by Hooke, Newton and others. In 1675

Newton described observations of colours in thin films of transparent material and from geometrical considerations was able to calculate effective film thickness and correlate this with the periodicity of observed colours. The films in question were typically of air or water in the space between two glass surfaces having different radii of curvature. The phenomena described are seen most conveniently as *Newton's rings*, where an optically flat glass surface is in contact with a convex glass surface having a large radius of curvature. Under monochromatic illumination in reflection, a series of concentric

coloured rings were observed, the radii being proportional to $\sqrt{1}$, $\sqrt{2}$, $\sqrt{3}$, etc. Although the wave nature of light and interference phenomena were not universally accepted at that time (especially by Newton) these observations, and others where the radii of the rings were seen to be larger for red than for violet light, established one of the most important properties of thin films. Since these early observations of thin film properties, they have been found to display not only characteristic and distinctive optical behaviour, but also mechanical, electrical and magnetic properties. These properties underpin many of their present-day industrial applications, such as antireflection coatings and optical filters, surface acoustic wave devices, electronic components (both discrete and integrated) such as resistors, capacitors, thin-film transistors and other active devices, magnetic data storage and superconductors.

Notwithstanding this wide range of properties and applications, it should be stressed that there is no watertight definition of the thickness below which a film becomes thin. The reason for this is primarily because different properties scale differently with thickness. *Eckertová* [29.1] has pointed out that it is permissible to say in general that the physical limit is determined by the thickness below which certain described anomalies appear, but that this differs for different physical phenomena. For instance a film which is optically thin may be of the order of the wavelength of light (≈ 500 nm) although some electrical properties are determined by the mean free path of conduction electrons (≈ 50 nm for metals) or by the thickness at which a given applied voltage produces an electric field which leads to high-field (nonlinear) behaviour. Generally speaking a thin film has a thickness of less than $1\text{ }\mu\text{m}$, and is deposited on the surface of a substrate by one of several distinct deposition methods [29.2]. In keeping with the remit of this Handbook, the emphasis in this chapter is on the electrical and optical properties of thin films, as these are clearly the most important in the field of electronic and optoelectronic materials. It is not the intention to dismiss thermal and mechanical properties as irrelevant in this sphere, but to stress that, although they may both also have an influence on the electrical and optical properties, they are both mainly dependent on the material itself. Mechanical properties of thin films have been reviewed by *Hoffman* [29.3] and also by *Campbell* [29.4]. However, the *structure* of thin films often has a crucial effect in determining other thin-film properties and, unlike the thermal and mechanical properties, is very largely determined by the film deposition method and conditions. For these reasons sections on both the struc-

ture and deposition methods of thin films are included in this review.

The basic structure of the chapter, following this brief introduction, covers the various aspects of thin films starting from their deposition methods, through their structures to their optical and electrical properties. There is no attempt to cover all of these aspects comprehensively; in a chapter such as this there is sufficient space to cover only the most important techniques and properties. In Sect. 29.1 the major deposition methods are described, such as physical processes like evaporation and sputtering, and chemical methods including chemical vapour deposition (CVD) and the Langmuir–Blodgett technique for molecular films. In Sect. 29.2 some examples of the crystalline structure of various types of films are examined. These can be relatively simple in the case of evaporated metal films, whereas compound films (such as compound semiconductors) can exhibit a variety of different structural forms depending on the deposition conditions and the thickness. Films prepared from larger molecules, such as the organic phthalocyanines having molecular weights of the order of 500, can also show several quite complex structures. In addition to the crystal structure per se, which is normally determined using X-ray, electron or neutron diffraction techniques, the morphologies of thin films often show interesting features, which are observed using electron microscopy or one of the newer scanning-probe imaging techniques such as scanning–tunnelling microscopy (STM) or atomic force microscopy (AFM). Section 29.3 covers the basic optical and electrical properties of thin films. The optical properties covered in Sect. 29.3.1 are largely the result of the electromagnetic wave nature of light, leading to the interference phenomena which were first observed in the 17th century. Films of suitable thickness and refractive index may be used in simple antireflection coatings, and quite complex multilayer film structures are used as optical filters. In Sect. 29.3.2 electrical properties are described. These include electrical conductivity (or resistivity) in relatively high-conductivity materials such as metals, where the conductivity depends not only on the carrier concentration and bulk mean free path, but is also modified by the effects of scattering at one or both of the film surfaces and internal grain boundaries. Lateral conduction in discontinuous (island) films where there are significant potential barriers between highly conducting regions has been investigated by *Neugebauer* and *Webb* [29.5], although this not discussed in the present work. This mechanism is reviewed elsewhere in the literature [29.1, 6]. Various nonlinear conduction pro-

cesses arise as the result of the high electric fields that can be applied across thin films; for example a film of thickness 100 nm having a potential difference of only 1 V applied between its surfaces, would experience a field of 1×10^7 V/m, which approaches the dielectric breakdown strength of many materials. Finally in Sect. 29.4 the main points of the discussion are summarised and some speculations concerning future trends are given.

Much of the technology associated with thin-film deposition and growth is explicitly excluded from this discussion. The measurement of film thickness, both during deposition (monitoring) or post-deposition, is not covered, although good accounts are given in the literature [29.7, 8]. Furthermore the condensation and growth mechanisms of thin films are also omitted [29.9, 10], although the structure and other properties are largely influenced by these. Many of the up-to-date materials properties and device applications are covered elsewhere in this volume. The reader is referred to Chapt. 14 for details of epitaxial growth techniques and to Chapt. 17 for structural characterisation. Optical properties and characterisation are discussed in Chapt. 3, while the corresponding materials and devices are addressed in Chapt. 37. Sensors and transducers, which are frequently based on various thin-film techniques (both inorganic and organic) are described in Chaps. 39 and 54. In the area of electronics, much current effort is in the areas of molecular electronics as described in Chapt. 53. Indeed the present author has predicted that many of the high-field conduction effects commonly observed in thin films will, in due course, also be detected in nanostructures in a suitably modified form [29.11].

Many of the topics covered in this chapter are explained in more detail in various books. The *Handbook of Thin Film Technology* [29.12], although published over three decades ago, has an enormous amount of detail concerning basic thin-film phenomena, most of which is still of interest today. Another very useful resource is the series of volumes originally entitled *Physics of Thin Films* and later simply *Thin Films* (recent volumes retitled *Thin Films and Nanotechnology*) published by Academic Press. Volume 25 of this series [29.13] contains an index for all the earlier volumes covering the years 1963–1998. Detailed references to several articles from both these sources are given in the references. Also of particular interest in the materials field is the *Handbook of Thin Film Materials* [29.14]. The reader is also referred to several other books covering the general field of thin films and technology [29.1, 15–19]. Other books covering more restricted areas of thin-film technology (e.g. optical properties, electronics) are referred to in the appropriate sections.

The SI system of units is used throughout, and where original work was presented in non-SI units, these have normally been converted. Other customary units, such as the electronvolt, are used where appropriate. The main emphasis in this chapter is on the physics of the deposition processes and the optical and electrical properties, and on their relationships with film structure, while keeping mathematical details to a minimum. As mentioned earlier, only the most important methods and properties are covered, although references to techniques omitted and fundamental factors are included to aid the acquisition of a deeper understanding, where this is required, and to appreciate the applications of thin films in those areas not covered elsewhere in this volume.

29.1 Deposition Methods

The deposition methods described in this section are those most commonly used. The physical methods of vacuum evaporation and sputtering are perhaps the most controllable, giving the best-quality films. In particular, molecular-beam epitaxy (MBE) is capable of providing very high-quality films. Examples of chemical methods, which are somewhat less flexible and limited to particular types of films, are also given. Perhaps the most important of these is CVD, but as will be seen the particular route chosen depends on the chemical properties of the material to be deposited.

29.1.1 Physical Deposition Methods

These could perhaps be described as physical vapour deposition (PVD) methods, since they involve direct deposition from the vapour phase, normally under reduced pressure. There are two main classes which are considered. The first of these is vacuum evaporation, which takes place at a low background pressure, i. e. high vacuum (HV) of the order of 10^{-4} Pa, although MBE systems normally operate at pressures significantly below this level. Secondly we have sputtering, which relies on the ejection of atoms or molecules from the surface

of a material target by a mechanism of momentum transfer involving an impact from (usually) ionised inert gas atoms such as argon; in order for this to occur a sufficient inert gas pressure is required to sustain a discharge, and thus the typical operating pressure is of the order of 1 Pa. Thus evaporated films are liable to have a higher intrinsic purity than sputtered films, which are prone to the incorporation of sputtering gas atoms. However, as these are generally inert there is often no deleterious effect on the film properties.

Vacuum Evaporation

During evaporation the evaporant is heated in vacuum by, for example, employing a high current through a refractory boat in which the evaporant is placed, or by using a high-current electron beam focussed on the evaporant. The purpose of this heating is to convert the condensed phase of the evaporant into the vapour phase, which then condenses elsewhere in the system in the form of a thin film. At a given absolute temperature T , there will in general be some of the evaporant in a condensed phase and some in the vapour phase. In thermodynamic equilibrium the Clausius–Clapeyron equation, relating the equilibrium vapour pressure p^* to the temperature T may be applied:

$$\frac{dp^*}{dT} = \frac{L}{T(\nu_g - \nu_c)}, \quad (29.1)$$

where L is the latent heat of vaporisation, representing the amount of energy required to convert a quantity of the condensed phase (liquid or solid) into the vapour phase at the given temperature. If L is expressed in J/mol, then ν_g and ν_c represent the molar volumes of the gaseous and condensed phases respectively. Although evaporation is not an equilibrium process, the importance of the equilibrium vapour pressure to evaporation is that maximum evaporation rates are dependent on the vapour pressure. Expressions for p^* may be obtained from (29.1) provided the temperature dependence of L is known. For the simplest assumption, where L is independent of T , we assume that the vapour obeys the ideal gas law. For one mole of the vapour

$$p^* \nu_g = RT, \quad (29.2)$$

where R is the universal gas constant. Further assuming that the volume of the vapour phase is very much greater than that of the condensed phase ($\nu_g \gg \nu_c$), combining the above two equations and integrating, an expression for p^* is obtained. This can be expressed in logarithmic form as

$$\log_{10} p^* = AT^{-1} + B, \quad (29.3)$$

where $A = -(L/R) \log_{10} e$ and B is a constant of integration. The constants A and B are tabulated by *Dushman and Lafferty* [29.21] for various metals using non-SI units, although in a slightly different form from the definitions of (29.3). In many cases such an expression for determining equilibrium vapour pressure is inadequate. The main reason for this is that the assumption of a constant latent heat L is invalid, particularly over a wide temperature range. When temperature variations in L are taken into account, (29.3) is modified by the addition of several additional terms, which may be determined from standard thermodynamic enthalpy and entropy data [29.22]. Tabulated values of coefficients for a logarithmic $p^* - T$ expression involving four coefficients are given by *Kubaschewski et al.* [29.23] for various elements and compounds over defined temperature ranges in both the liquid and solid phases. Even more helpful are direct tabulated and graphical data of the $p^* - T$ relationship in the case of various solid and liquid elements [29.20, 22, 24]. Examples of the $p^* - T$ dependence are given in Fig. 29.1 for several elements using the data of *Honig and Kramer* [29.20]. From this figure it is clear that p^* increases very rapidly with temperature, and that most metals require temperatures of up to about 1500 K to attain significant vapour pressures; however, the refractory metals (Pt, Ta, W, etc.) require far higher temperatures for evaporation and thus they are not normally deposited by this method, although they are very useful materials for evaporation boat ma-

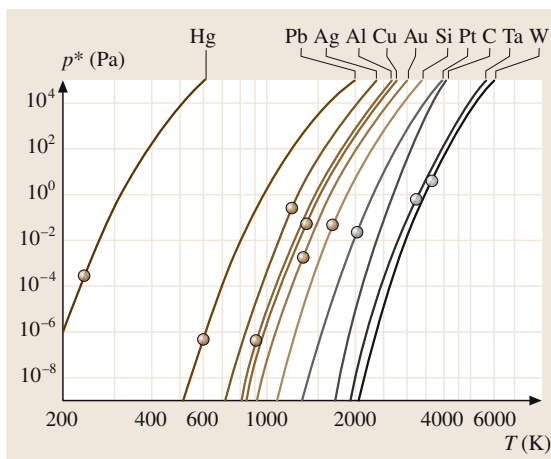


Fig. 29.1 Experimental dependence of the equilibrium vapour pressure p^* on absolute temperature T for several elements, derived from the compilation of *Honig and Kramer* [29.20]. The point (•) on each curve indicates the melting point, except for that of carbon, which sublimes

terials, which require very low vapour pressures and high melting points.

It is a well-known result from the kinetic theory of gases, that the molecules of a gas at a pressure p impinge on surfaces at a rate

$$N_i = \frac{p}{(2\pi mk_B T)^{1/2}} \quad (29.4)$$

where N_i ($\text{m}^{-2}\text{s}^{-1}$) is termed the impingement rate, m is the mass of a gas molecule, k is Boltzmann's constant and T is the absolute temperature. In principle this equation applies equally well to residual gas molecules (N_2 and O_2) in an evaporation chamber, as well as to evaporant molecules in the vapour phase. It was noted previously that the equilibrium vapour pressure of a material largely determines the maximum evaporation rate. Hertz [29.25] observed for the case of mercury that evaporation rates were proportional to the difference between the equilibrium vapour pressure p^* and the reverse hydrostatic pressure p_h exerted at the surface of the evaporant. The *maximum* evaporation rate occurs when $p_h = 0$, and is equal to the impingement rate a vapour of the evaporant would exert at its equilibrium vapour pressure. This is given by an expression similar to (29.4), with pressure p^* , evaporation temperature T_e and mass of evaporant molecule m_e . This rate is rarely achieved, not only due to the hydrostatic pressure p_h , but also to a molecular reflection phenomenon proposed by Knudsen [29.26], which takes place at the evaporant surface. It was proposed that only a certain fraction α_v (the evaporation coefficient) of molecules make the transition from the condensed to the vapour phase. The net molecular evaporation rate, N_e , is then given by the Hertz-Knudsen equation

$$N_e = \frac{\alpha_v (p^* - p_h)}{(2\pi m_e k_B T_e)^{1/2}} \quad (29.5)$$

Under equilibrium conditions $p_h = p^*$ and there is no net evaporation; however, under nonequilibrium conditions, especially when $p_h \ll p^*$, the evaporation rate can be considerable. Values of the coefficient α_v have been tabulated in the literature for a selection of different elements and compounds [29.27].

Equipped with vapour pressure and evaporation coefficient data, it is straightforward to estimate the evaporation rate using (29.5). The molecular deposition rate N_R represents the number of evaporated molecules deposited on the substrate per unit area per second. It is clearly proportional to the evaporation rate N_e , depending on the geometry of the evaporation/deposition system. N_R is simply related to the thickness deposition

rate R (m/s) by

$$N_R = \frac{\rho R}{m_e} \quad (29.6)$$

where ρ is the density of the depositing film. An indication of the film quality is given by the impingement ratio K , representing the ratio of the rate at which ambient gas molecules impinge on the substrate to the rate of evaporated molecules depositing on the substrate. Hence

$$K = \frac{N_i}{N_R} \quad (29.7)$$

From (29.4) and (29.6) it is clear that $K \propto p/R$ for a given material. In order to reduce contamination by gas incorporation into the film, it is therefore necessary to operate at a low background gas pressure with a relatively high deposition rate. A simple calculation for copper shows that for $R = 1 \text{ nm/s}$, $N_R \approx 8.5 \times 10^{19} \text{ m}^{-2} \text{ s}^{-1}$. If we require $K < 10^{-3}$ for impinging oxygen molecules then the maximum value of N_i allowed is $8.5 \times 10^{16} \text{ m}^{-2} \text{ s}^{-1}$, which from (29.4) corresponds at room temperature to a pressure of about $3 \times 10^{-6} \text{ Pa}$. Such a pressure is in the range of ultra-high vacuum (UHV), and is not attainable in normal vacuum deposition systems. Increasing the deposition rate is the normal method of lowering K without moving into UHV, as is used in MBE. This may in principle be achieved either by increasing the evaporation rate by increasing T_e to raise p^* , or by increasing the proportionality constant between N_e and N_R by shortening the distance between the evaporant and the substrate. A useful table of values of K for various combinations of chamber pressure and deposition rate is given by Eckertová [29.1].

There is in principle a second limitation on the level of the background pressure in a vacuum system which is also derived from kinetic theory. By using a Maxwell-Boltzmann distribution function for the molecular velocities in a gas, it can be shown that the mean free path between collisions in the gas λ is given by [29.22]

$$\lambda = \frac{k_B T}{p\pi\delta^2\sqrt{2}} \quad (29.8)$$

where δ is the effective molecular diameter of a gas molecule. Values of δ are tabulated in the literature [29.21], and range from about 0.22 nm (He) to 0.49 nm (Xe). From (29.8) λ is inversely proportional to the pressure p , and this general relationship is shown for He, O_2 and Xe in Fig. 29.2. Since the values of δ do not vary greatly for different gases, the λ values for all common gases fall

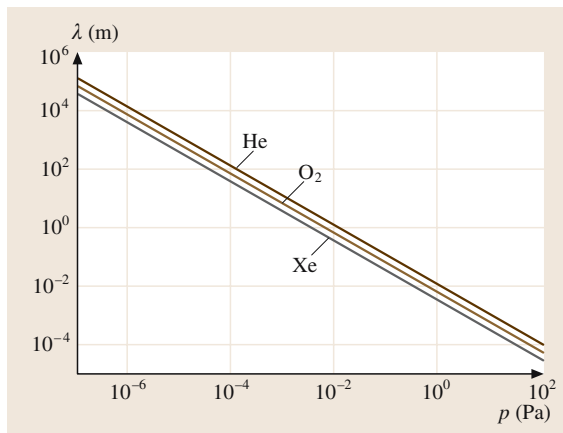


Fig. 29.2 The dependence of mean free path λ on pressure p for helium, oxygen and xenon at 300 K. Mean free paths are calculated from (29.8) using values of molecular diameter δ from Dushman and Lafferty [29.21]

within a narrow range, as illustrated in the figure. The proportion of molecules $F(x)$ which have not suffered an intermolecular collision as a function of distance x since their last collision is given by

$$F(x) = \exp(-x/\lambda). \quad (29.9)$$

From this expression it is evident that for a deposition chamber size of about $\lambda/10$, less than 10% of molecules suffer a collision before hitting the walls of the chamber or the substrate, e.g. in a large vacuum chamber of dimensions about 1 m where the mean free path is about 10 m. From Fig. 29.2 a pressure of 10^{-4} Pa is sufficiently low to ensure that $\lambda > 10$ m. In evaporation processes we require that the evaporant molecules are not significantly dispersed by collisions with gas molecules en route to the substrate, and since the mean free path is relatively insensitive to the molecular diameter, λ is a suitable criterion to determine this. Thus smaller deposition chambers will not require as low a residual gas pressure as larger ones.

The geometry of the evaporation source and its positioning with respect to the substrate is an important consideration in determining the uniformity of a thin-film deposit. Expressions for the film thickness variation over the substrate for various types of source have been derived by Holland and Steckelmacher [29.28]. Traditionally the *point source* and the *small surface source* are used. It is assumed that the evaporation source, of either type, is located a distance h below the centre of a flat substrate, as shown in the inset to Fig. 29.3. It is obvious that the film thickness deposited directly above

the source at the point O will be thicker than elsewhere, since this is the closest point of the source to the substrate. For the case of the point source, molecules are emitted uniformly in all directions, and the ratio of the film thickness d deposited at the point P a distance l from O to the thickness d_0 deposited at O is given by

$$\frac{d}{d_0} = \frac{1}{[1 + (l/h)^2]^{3/2}}. \quad (29.10)$$

For a small surface source, the emission follows a cosine law, and the thickness distribution is given by

$$\frac{d}{d_0} = \frac{1}{[1 + (l/h)^2]^2}. \quad (29.11)$$

Plots of these two functions are shown on Fig. 29.3 for the point source P and the small surface source S. For both sources, there is a rapid diminution in the thickness ratio d/d_0 , which falls off to 0.985 and 0.98 respectively for the point and surface sources at a value of $l/h = 0.1$. When the substrate is relatively small, the thickness uniformity is thus better than 2%. Uniformity can be enhanced, for these two small sources, by increasing the source–substrate distance h , but at the expense of consuming more evaporant material. It has also been shown that a substantially uniform deposit will be obtained for a point source if the source is placed at the centre of

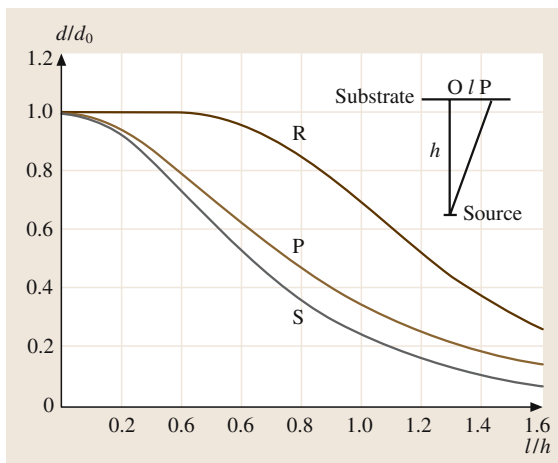


Fig. 29.3 Film thickness profiles calculated for evaporation from various sources onto a flat parallel substrate. Thickness ratio d/d_0 is given as a function of the ratio l/h , as shown in the inset. Sources include the small surface source (S), the point source (P) and a thin ring source (R) of radius b , where $b/h = 0.75$. Profiles are calculated, respectively, from (29.11), (29.10) and from an expression given by Glang [29.22]

a sphere and the substrates placed on the surface. For a surface source a uniform thickness distribution is obtained if both the source and the substrates are placed on the surface of a sphere [29.28].

Other systems for obtaining uniform thickness are based on the use of either extended sources, or substrates rotating around a central axis with respect to an off-centre source, such that variations in the thickness average out. *Holland* and *Steckelmacher* [29.28] have addressed ring sources in addition to strip and cylindrical sources evaporating onto a plane-parallel substrate. The ring source has also been treated by *von Hippel* [29.29] and the small surface source evaporating onto a rotating plane parallel substrate has been discussed by *Macleod* [29.30]. Expressions for ring and large surface sources are given by *Glang* [29.22]. A ring source can be considered to be a narrow ring of radius b , which is usually placed parallel to a flat substrate. This introduces an additional controllable variable which influences the thickness distribution. Thickness distributions may be obtained by calculating the distribution arising from an element of the ring source and then summing over all elements in the ring. For small b/h the thickness distribution reduces to that of the small surface source. For b/h larger than about unity the uniformity is also poor as the maximum in the distribution moves away from the central axis towards the source radius. However, for intermediate ranges, in particular b/h of about 0.7–0.8 there is considerable uniformity in thickness up to l/h of about 0.6. This is also illustrated in Fig. 29.3, where the curve R shows the thickness distribution for a narrow ring source with $b/h = 0.75$, calculated from the formula given by *Glang* [29.22]. Extended disc source thickness distributions may be obtained by integrating the thickness distribution derived for a narrow ring source from radius zero to radius b , but in general the thickness uniformity is not as good as for ring sources. Expressions for ring source of finite width can be determined by subtracting the disc source distribution for the inner radius from that of the outer radius. Diagrams illustrating the thickness distributions for several sets of ring and disc sources are given by *Glang* [29.22]. *Behrndt* [29.7] gives further details and references for such sources, and a discussion of the results of using combinations of several rod-like sources. Further attempts to minimise thickness variations have resulted in the evolution of various *planetary* deposition systems, so called because the rotating substrate plane is itself rotated around a vertical axis. A typical example of such a system and its performance is given in the literature [29.31].

Before describing the practical details of the evaporation process, it should be mentioned that, whatever type of evaporation process is used, some consideration needs to be given to the substrate. There are three basic types of substrate commonly used: glasses, single crystals and polycrystalline ceramics. Ceramics are particularly useful for high-temperature deposition processes. Although the substrate material may be predetermined (e.g. silicon wafers for microelectronics, optical glass for coatings on lenses, etc.), there are frequently factors which can be selected to increase the probability of obtaining good-quality films. The substrate planarity is important, in order to reduce the effect of defects. Electrical resistivity should in general be high in order to minimise surface leakage currents. Thermal properties include a consideration of the coefficient of expansion, which should be comparable to that of the deposited film to reduce the probability of thermally induced stresses in the films. Thermal conductivity should be high for films whose applications will entail significant power dissipation. A full description of thin-film substrates and the associated problems of substrate cleaning are discussed in detail by *Brown* [29.32].

Resistive evaporation is the most commonly used evaporation technique, although it is unfortunately not suitable for certain applications. It entails the resistive heating of an evaporation source, by passing a high current either through the source itself or through an adjacent heater. There are a large number of practical evaporation sources for various applications which are described in the literature [29.1, 22, 28]. The main requirement for any evaporation source is that the source itself will not emit its own vapour at the operating temperature. If metallic sources are used, the equilibrium vapour pressure of the metal must be negligible in comparison to that of the evaporant. A glance at Fig. 29.1 shows that the refractory metals (such as Pt, Ta and W) have equilibrium vapour pressures very low compared to about 1 Pa at temperatures up to about 2000 K, and are therefore suitable for evaporation of many materials. Carbon and oxides of aluminium or magnesium may also be used in evaporation, although owing to their low electrical conductivity refractory metal heaters are often used in conjunction with these. However, at elevated temperatures the oxides themselves may dissociate, with the liberated oxygen reacting with the evaporant to form its own oxide. For this reason the refractory metals cannot themselves be evaporated from oxide crucibles. Furthermore carbon may react with metals to form carbides well below 2000 K. Having chosen a suitable evaporation source, the vacuum deposition

chamber is pumped down to the required background pressure and deposition may commence. The temperature of the source is increased such that the vapour pressure increases to a value consistent with a reasonable evaporation and hence deposition rate. The substrate is normally shielded using a metallic shutter until the required deposition rate is established, when the shutter is removed. The deposition rate is usually monitored by one of several methods [29.7, 22]. The quartz crystal method entails depositing the evaporant material on a vibrating quartz crystal, whose frequency varies linearly with the thickness of material deposited [29.33, 34]; the deposition rate is obtained by electronically differentiating the thickness signal with respect to time. Optical thickness monitoring methods are also used, particularly in applications where films are required for optical purposes, and require films having an optical thickness of multiples of $\lambda/4$, where λ is the wavelength for which the optical film is designed [29.7, 22]. These thickness multiples correspond to maxima and minima in reflected light intensity, which can easily be monitored during deposition. When the required film thickness is attained, a shutter is again used to curtail further film deposition and the evaporation process is stopped. Sequences of several film layers may be deposited by using a suitable sequential masking system [29.35].

Apart from simple vacuum evaporation there are also several other types of evaporation, which entail using alternative methods to heat the evaporant, or cater for various problems encountered in the deposition of alloys and compounds. In electron-beam evaporation the evaporant material is placed in a ceramic hearth, which is bombarded by high-energy electrons. The main advantage of this method is that considerably higher temperatures of above 3000 K may be obtained than with resistive evaporation methods. This allows evaporation of the refractory metals and also elements such as boron and carbon (which sublimes). Other advantages are the evaporation of reactive metals such as Al, which are prone to contamination due to alloying, and situations where high purity is required.

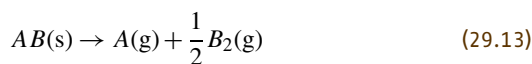
For non-elemental (multicomponent) films there are several methods which may be employed. The main problem with such films is that, because of differences between the vapour pressures of the evaporant components, the Hertz–Knudsen expression of (29.5) predicts that the compound will evaporate noncongruently, i.e. with different evaporation rates for each element, and the deposited film will not in general replicate the composition of the evaporant. The problem is compounded

because the proportion of each component present will therefore vary throughout the evaporation process, and the overall composition of the deposited films will vary throughout its thickness. In the evaporation of alloys, each component establishes its own vapour pressure at a given temperature, which may be predicted by *Raoult's law* for an ideal solution. This states that the vapour pressure established by each individual component is proportional to the mole fraction of the component present. Under these circumstances it is relatively simple to adjust the proportions of the alloy components to achieve the required film composition. However, complications occur in nonideal solutions, where the interactions between different types of atoms vary and an empirical parameter known as the activity coefficient, a , is defined. Knowing this it is possible to determine the vapour pressure for the component and hence the evaporation and deposition rates. Activity coefficients for various important systems, such as nichrome (a Ni–Cr alloy), are given in the literature [29.36]. However, as noted previously the mole fractions present in the evaporant are time-dependent and the composition of the films will vary during deposition. A parameter K has been determined, which if maintained close to unity should result in congruent evaporation throughout the entire deposition process. Further details are given by *Zinsmeister* [29.37]. Additional comments on the evaporation of both nichrome and permalloy alloys are given by *Glang* [29.22].

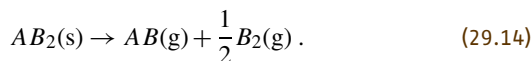
On evaporation of compounds, a wealth of chemical reactions can occur, which leads to a variety of different evaporation mechanisms. Frequently more than one reaction may operate simultaneously. Four simple generic reactions have been listed by *Glang* [29.22], which include many of the materials which can be successfully evaporated. We consider only compounds AB composed of the elements A and B , and denote the phase of the material by s (solid), l (liquid) or g (gas). Some compounds (solids or liquids) do not dissociate on evaporation and the reaction can be expressed simply as



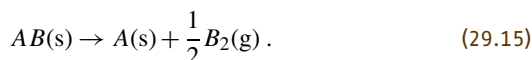
There is then only the single vapour phase $AB(g)$ produced. Other compounds dissociate on evaporation into two different vapour species. One reaction which is typical of the evaporation of chalcogenides and some simple oxides is given by



whereas the group IV dioxides tend to dissociate according to the reaction



In compounds of elements whose volatilities are significantly different, frequently one element enters the vapour phase while the other remains in the solid (or possibly liquid) phase. The compound decomposes into its different elements, with the simplest reaction of this type being



Examples of materials evaporating without dissociation following (29.12) are the simple oxides GeO, SiO, SnO, the fluorides CaF₂ and MgF₂, and PbS. In general the evaporation of these materials can therefore be considered as analogous to those of the elements, and where available simple p^*-T data give the vapour pressure. Materials which follow (29.13) are primarily the II–VI chalcogenides, in particular the cadmium compounds CdS, CdSe and CdTe, and several simple oxides such as BaO, BeO, CaO, MgO, NiO and SrO. A full review of the electrical and structural properties of the cadmium chalcogenides is given by *Gould* [29.38]. Equation (29.14) is followed by many of the group IV dioxides, such as SiO₂, SnO₂, TiO₂ and ZrO₂, which dissociate into their lower oxides and molecular oxygen. Decomposition, as described by (29.15) is followed by most of the metallic borides, carbides and nitrides [29.22]. Other modes of decomposition, such as those for the technologically important III–V compounds, are also described in this reference.

Provided the relevant reaction for the evaporation process is known [e.g. as given by Eqs. (29.12–29.15)] it is possible to determine the equilibrium vapour pressure using standard thermodynamic tables. The law of mass action may be used to determine the equilibrium state of the chemical reaction, and for evaporation, the pressures of the various phases present contribute to a temperature-dependent equilibrium constant K_p , which is given by

$$K_p = \prod_i (P_i)^{v_i} \quad (29.16)$$

where P_i represents the vapour pressure for each of the various components of the vapour expressed in standard atmospheres and the v_i are the molecular coefficients. K_p is related to the thermodynamic parameter ΔG^0 , the standard change in Gibbs free energy per mole for the reaction, by the expression

$$\Delta G^0 = -RT \ln K_p, \quad (29.17)$$

where R is the universal gas constant. Free-energy data is tabulated for many materials in the literature [29.23, 39] and the use of (29.16) and (29.17) allows realistic estimates of the equilibrium vapour pressures to be determined and the evaporation rates estimated. Further discussions are given by *Glang* [29.22]. Nevertheless, although the theoretical framework is in place for the controlled evaporation of compounds, the exact type of dissociation and decomposition are not known in all cases, and for compounds consisting of three or more elements this problem is compounded. Some materials are therefore not amenable to direct evaporation from a single source.

For these cases various special evaporation techniques have been developed, all of which aim to replicate the composition of the evaporant in the deposited film, by ensuring that the different molecules arrive at the substrate in the desired proportions and at a constant rate. Reactive evaporation is a process whereby one component of the desired film is evaporated resistively in the normal manner, while the other component is present in the evaporation chamber in the form of a gas. This method is most commonly used where oxide films are required and oxygen gas present in the chamber combines with a metal at the substrate to form the oxide. The impingement ratio K , given by (29.7) is optimised so that evaporant and gas molecules reach the substrate in a predetermined ratio. Oxides of the common metals have been deposited by this method, including those of Al, Cr, Cu and Fe; additionally tantalum and titanium oxides have been prepared for dielectric applications. Reactive evaporation has also been used for compounds other than oxides. For instance, stoichiometric CdS films have been obtained by the evaporation of CdS in a sulphur vapour [29.40]. The sulphur vapour has the effect of discouraging the appearance of nonstoichiometric Cd-rich films which occur when CdS is evaporated alone. Nitrides may also be deposited in some cases by the use of a nitrogen atmosphere. The utility of this technique has been demonstrated in the reactive evaporation of silicon oxide, SiO_x, by *Timson* and *Hogarth* [29.41]. The evaporant is a mixture of the monoxide, the dioxide and possibly also free silicon. A sequence of films, varying in composition between SiO and SiO₂ as determined by electron spin resonance studies were obtained, depending on the ratio of the deposition rate to the oxygen pressure R/p . A useful nomogram for use in reactive evaporation is given by *Glang* [29.22].

Co-evaporation, sometimes known as the three-temperature technique, involves the simultaneous evaporation of two or more materials from separate evap-

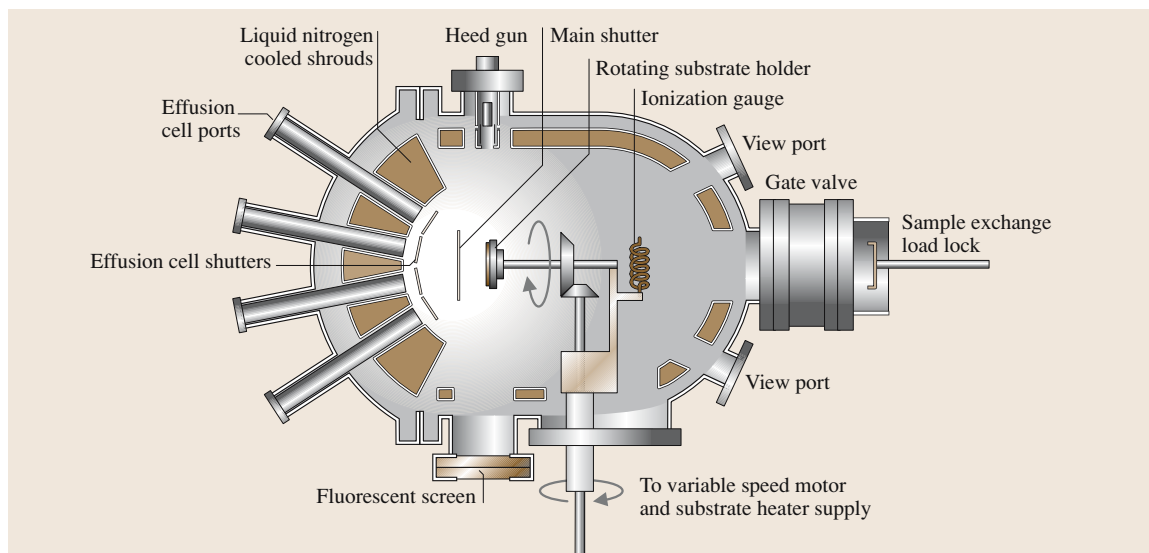


Fig. 29.4 Schematic diagram of a molecular-beam epitaxy (MBE) system for the deposition of III–V semiconductor compounds. After [29.42] with permission from Elsevier Science

oration sources. The essential feature of this process is that the rate of molecular arrival at the substrate may be controlled by individually controlling the temperatures of the evaporation sources. A major advantage of this method is that materials with widely different vapour pressures at a given temperature may be co-deposited by increasing the temperature of the less-volatile material. This method has found applications in the deposition of II–VI compounds such as CdS and CdSe, and also the III–V compounds such as AlSb, GaAs, InAs and InSb. It is also used for the deposition of alloy films.

Flash evaporation is another technique used for deposition of layers whose components have widely different vapour pressures. The evaporation is constrained so that only very small amounts of the compound completely vaporise almost instantaneously. The various components normally evaporate noncongruently, but because only very small quantities are evaporated rapidly, any inhomogeneity in the deposited films is limited to a few atomic layers. The process is repeated by evaporating further small quantities of evaporant. Several alternative arrangements have been developed to evaporate small particles, by continuously feeding powdered evaporant from a hopper into a heated evaporation source. The rate of delivery needs to be controlled such that at any given time there will be a number of evaporant particles in the source at various stages of the evaporation process. This ensures that the vapour emerging from the source contains contributions from many individ-

ual particles, resulting in an overall vapour composition which approximates to that of the evaporant powder. The method has been used particularly for III–V compound semiconductors such as GaAs.

Finally, the method of molecular-beam epitaxy (MBE) should be included here as a particularly sophisticated example of evaporation (or co-evaporation). In MBE the background pressure is considerably lower than in the case of the simpler evaporation techniques, i. e. well into the UHV range. This effectively eliminates the effects of background gas impingement, drastically improving the purity of the films. Several sources (effusion ovens) may be used simultaneously, each of which is separately controlled. These permit the deposition of compounds and precisely controlled doping of semiconductors. Sophisticated monitoring and isolation techniques are employed, and deposition sequences are usually under computer control. MBE is used particularly for the preparation of III–V compounds [29.42, 43] and silicon [29.44]. Figure 29.4 illustrates a typical MBE system for use in the deposition of III–V compounds [29.42]. A full description of MBE, together with other epitaxial growth techniques is given elsewhere in this volume in Chapt. 14.

Sputtering

The technique of sputtering in its various forms is of secondary importance only to evaporation for the deposition of thin films, offering certain advantages for the de-

position of high-melting-point and dielectric materials. Sputtering is essentially the removal of particles from the surface of a *target* of the deposition material, by the action of incident energetic particles (normally positive ions). It is generally considered that the sputtering mechanism is the result of the transfer of momentum from the incident particles to the target atoms, some of which become dislodged. Energy acquired by other atoms interacting with the dislodged atoms can be sufficient to overcome the surface binding energy potential barrier. The net effect of a large number of atomic interactions is for a certain proportion of the target to be released (or sputtered) from the surface. Sputtered atoms typically have energies much higher than their evaporated counterparts. For low-energy sputtering with incident energy 1 keV, an ejected particle may have an energy of 10 eV, since typically 1% of the incident energy is transferred to each sputtered atom. In contrast, the thermal (kT) energy of an atom evaporated from a source at 2000 K is less than 0.2 eV. Such considerations have implications regarding the sticking coefficients for deposited materials (i.e. the probability that an atom reaching the substrate condenses on it), which is generally higher for higher-energy atoms.

Sputtering is limited by a threshold energy, below which it will not occur for a given ion/target combination. Threshold energies for a wide range of incident ions and target materials are of the order of 20–30 eV; experimental values for some common ion/target combinations are given by *Wehner and Anderson* [29.45]. These workers have pointed out that the threshold energy appears to be approximately four times the latent heat of sublimation per atom, equivalent to the surface binding energy. However, for practical sputtering considerations, incident energies are normally at least a few hundred electronvolts, well above the threshold energy. The efficiency of sputtering, measured by the sputtering yield, is also very low at low incident ion energies.

The sputtering yield is the most important sputtering parameter, and its variation with the target material and the incident ion energies is of major concern in both theories of sputtering and in practical sputtering deposition techniques. Sputtering yield $S(E)$ is defined as the ratio of the number of sputtered atoms to the number of incident ions. Hence

$$S(E) = \frac{N_s}{N_{\text{ion}}} , \quad (29.18)$$

where N_{ion} is the impingement rate of incident ions and N_s is the sputtering rate of the target. Its primary

variation is with the incident ion energy E . However, it also depends on the atomic properties of the incident ions and the target atoms. In the following discussion Z_1 and m_1 refer to the atomic number and mass of the incident ions, and Z_2 and m_2 refer to the corresponding quantities for the target material. The dependence of sputtering yield on the atomic number of the incident ions is quite striking, as shown in Fig. 29.5 [29.46]. This shows the variations of $S(E)$ for 45-keV incident ions onto targets of silver, copper and tantalum. Maxima in $S(E)$ occur corresponding to the noble gases Ne, Ar, Kr, and Xe. This dependence is one reason why the noble gases are most commonly used as the incident ion species, the second being that their inclusion into deposited films is unlikely to cause significant problems, due to their chemically inactive nature.

A number of theories have been developed to predict the dependence of sputtering yield $S(E)$ on the incident beam energy E . *Keywell* [29.47] assumed that the incident ions interacted with the sputtering target as if they were hard spheres, with the incident ion energy decreasing exponentially with the number of collisions. A discussion of this and other early models is given in the literature [29.45, 48]. More recent models [29.49, 50] have taken account of the scattering of incident ions by the atomic nucleus, and screening effects due to the electron cloud are also included, based on the Thomas–Fermi potential. This predicts a characteristic screening length a , which depends on the atomic numbers of the incident and target atoms, Z_1 and Z_2 , respectively. The screening length is given by

$$a = \frac{a_0}{(Z_1^{2/3} + Z_2^{2/3})^{1/2}} , \quad (29.19)$$

where $a_0 = 5.29 \times 10^{-11}$ m is the Bohr radius of the ground state of the hydrogen atom. Another parameter which is often invoked in sputtering models is the energy transfer coefficient, γ . For an incident ion of mass m_1 and energy E , impacting on a stationary target atom of mass m_2 , the maximum energy E_{max} the target atom will attain is given by

$$E_{\text{max}} = \gamma E , \quad (29.20)$$

where

$$\gamma = \frac{4m_1m_2}{(m_1 + m_2)^2} . \quad (29.21)$$

In most collisions the incident ion will lose some energy to the target lattice, resulting in the acquisition of an energy less than E_{max} by the target atom. γ has a maximum value of unity for $m_1 = m_2$, but

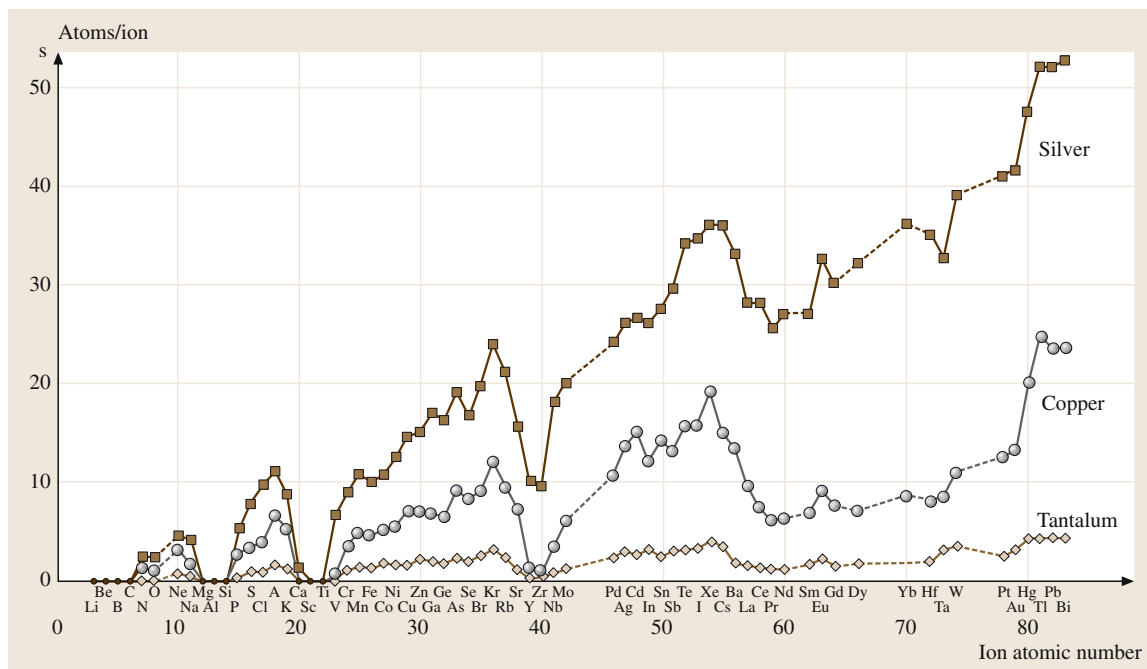


Fig. 29.5 Sputtering ratios of silver, copper and tantalum targets for different ions of energy 45 keV. After [29.46] with permission from Elsevier Science

this is neither a sufficient nor a practical condition for obtaining high sputtering yields owing to the impracticability of having identical sputtering ions and target material as well as the striking periodic dependence of $S(E)$ on Z_1 noted earlier. Two well-known sputtering models are now outlined, both of which have been reasonably successful in predicting sputtering yields.

In the model of *Pease* [29.49], which is also outlined elsewhere in the literature [29.51], ions are considered to interact with the target atoms near to the surface, which if displaced from their lattice sites are termed *primary knock-ons*. The mean energy of primary knock-ons before further interactions is denoted by \bar{E} . The threshold energy for displacement of a target atom, either by an incident ion or by a primary knock-on is denoted by E_d . Collisions with stationary atoms result in slowing down of displaced atoms, and in some cases a reversal of momentum and escape from the target surface, providing they still retain energy greater than E_s , the surface binding energy per atom. The number of atomic layers which contribute to the sputtering process is established by considering the average number of collisions n that a primary knock-on of energy \bar{E} will make before its kinetic energy is reduced to less than that re-

quired to leave the target surface, i.e. E_s . This is given by $2^n = \bar{E}/E_s$ or

$$n = \frac{\ln(\bar{E}/E_s)}{\ln 2}. \quad (29.22)$$

While engaged in n collisions the primary knock-ons diffuse on average $n^{1/2}$ atomic layers toward the surface, and thus $(1 + n^{1/2})$ layers are involved in sputtering of material liberated by primary knock-ons only. The total number of atomic layers involved in sputtering is also equal to this expression. We also have to consider the probability of interaction in a given atomic layer, and denote the number of atoms per unit volume in the target by N . If these atoms are considered to be arranged in uniform crystallographic layers in three dimensions, the atoms in each layer will have an area density of $N^{2/3}$. If σ_p denotes the cross section for interaction between incident ions and target atoms, the probability of interaction in a given atomic layer is given by the product of the number of atoms per unit area and their interaction cross sections, or $\sigma_p N^{2/3}$. The total probability of producing a primary knock-on which can contribute to the sputtering process is then given by the product of the probability of interaction in a given atomic layer and the number of layers involved in sputtering, i.e.

$\sigma_p N^{2/3} (1 + n^{1/2})$. An estimate of the number of atoms displaced per primary knock-on is $\bar{E}/2E_d$ [29.52] of which half, or $\bar{E}/4E_d$, will be directed towards the surface. The total sputtering yield $S(E)$ is then given by the product of the probability of producing primary knock-ons and the number of displaced atoms produced by each. Hence

$$S(E) = \sigma_p \bar{E} N^{2/3} \frac{1}{4E_d} \left[1 + \left(\frac{\ln(E/E_s)}{\ln 2} \right)^{2/3} \right] \quad (29.23)$$

where the expression for n in (29.22) has been substituted.

In (29.23) for the sputtering yield, N , E_d and E_s are constants depending on the target material and E_d is normally determined from radiation damage experiments. However, σ_p and \bar{E} are energy-dependent variables, which depend on the atomic numbers Z_1 and Z_2 , the Thomas–Fermi screening length a , and the maximum target-atom energy E_{\max} . It is not appropriate to give these expressions here, but it is worth mentioning that different expressions are applicable in different incident ion energy ranges, defined by $E < L_A$, $L_A < E < L_B$, and $L_B < E$. Further details concerning the threshold energies L_A and L_B are given in [29.49, 51]. Calculations of sputtering yield predicted by this model are reasonably consistent with experiment, showing a sharp increase above a certain threshold energy which is determined by E_s , a broad maximum and a relatively slow decrease at higher energies [29.49].

The model of *Sigmund* [29.50] represents a comprehensive theory of sputtering, particularly for single-element solids, although this has led to a certain degree of complexity in its general form. The model has been revised and extended [29.48] and tested against a considerable amount of data produced by many different workers [29.53]. The emphasis in this section is to briefly outline the model, to provide a basis for the calculation of $S(E)$ at normal incidence for low- and high-energy sputtering, and to give useful references to the required data. Various approximations made and some of the conditions of validity are omitted. Being a multiple-collision process, sputtering may be treated by using the formalism of transport theory, and in the Sigmund model the Boltzmann transport equation is solved. The sputtering yield at normal incidence is given by

$$S(E) = A \alpha N S_n(E), \quad (29.24)$$

where A is termed the material factor, and depends only on the target material. α is a dimensionless factor, effectively depending only on the ratio of the target atom

to incident ion masses m_2/m_1 . The factor α is given for various different conditions in the literature [29.48, 50]. As in the model of *Pease* [29.49], N represents the number of atoms per unit volume in the target. $S_n(E)$ is the nuclear scattering cross section, depending on the interaction probability between the incident ions and the target atoms at a given energy.

The material factor A arises in the sputtering yield expression as a result of integrating the number of sputtered atoms emitted in all directions with all possible energies. According to Sigmund, the material factor

$$A = \frac{3}{4\pi^2} \frac{1}{NC_0 E_s}, \quad (29.25)$$

where $C_0 = \frac{1}{2}\pi\lambda_0 a_{\text{BM}}^2$, λ_0 is a constant approximately equal to 24, and $a_{\text{BM}} = 2.19 \times 10^{-11}$ m is the Born–Mayer potential characteristic screening radius. E_s represents the surface barrier energy, as in the model of Pease. Making these substitutions into (29.24) gives a more useful formula for the sputtering yield [29.45] as

$$S(E) = \frac{1}{16} \frac{\alpha S_n(E)}{\pi^3 a_{\text{BM}}^2 E_s} \quad (29.26)$$

which applies both for low- and high-energy sputtering. An expression for E^* , the boundary between low-energy and high-energy sputtering, is given in the literature [29.50], and is typically a few hundred electronvolts, depending on the atomic masses and atomic numbers of the incident ions and the target atoms. For low-energy sputtering ($E < E^*$) the nuclear scattering cross section can be approximated by

$$S_n(E) = C_0 E_{\max} = \frac{1}{2}\pi\lambda_0 a_{\text{BM}}^2 E_{\max}, \quad (29.27)$$

where $E_{\max} = \gamma E$ from (29.20). Substituting into (29.26) yields

$$S(E) = \frac{3}{4} \frac{\alpha \gamma E}{\pi^2 E_s}. \quad (29.28)$$

Hence for low-energy sputtering $S(E)$ is directly proportional to the sputtering energy, where α and γ effectively depend only on the ratio m_2/m_1 and E_s is a constant for the target material.

For higher incident ion energies ($E > E^*$) the Thomas–Fermi, rather than the Born–Mayer interaction potential is applicable, and a simple approximation for $S_n(E)$ is not available. It is therefore necessary to determine its value from tables or graphical data, calculating $S(E)$ directly from (29.26). Incident ion energies E and nuclear scattering cross sections $S_n(E)$ may be expressed in terms of reduced or Thomas–Fermi variables ε and $s_n(\varepsilon)$ [29.48, 50]. $s_n(\varepsilon)$ is a universal function

of ε , depending on the details of the screened Coulomb function used [29.48]. The detailed dependence of $s_n(\varepsilon)$ on ε utilised in the Sigmund model is that given by Lindhard et al. [29.54]. For ease of computation E_{TF} , the Thomas–Fermi energy unit has been calculated for various incident ion/target combinations [29.48]. E_{TF} represents the ratio E/ε , and thus ε and hence $s_n(\varepsilon)$ may be obtained. $S_n(E)$ may then be determined from the reduced value.

Agreement between predicted and measured sputtering yields as a function of incident ion energy are good. Andersen and Bay [29.53] have reviewed sputtering yield measurements at normal incidence for a considerable number of incident ion and target species, and generally concluded that sputtering yields for high-yield materials (e.g. Ag, Au, Cu, Zn) are predicted well by the theory, while for low-yield materials (e.g. Nb, Ta, Ti, W) it was over-estimated by a factor of up to three. A fuller discussion, to which the reader is referred, shows that the model breaks down for certain combinations of Z_1 , Z_2 and E due to the existence of inelastic collisions, non-isotropic and nonlinear effects (spikes) in the collision cascade and unallowed-for surface effects. The regions of validity of the model are neatly presented in terms of a three-dimensional (Z_1 , Z_2 , E) space, and illustrate the fact that (29.26) is an amazingly good prediction for a large number of (Z_1 , Z_2 , E) combinations of practical interest.

Sputtering yields may be determined experimentally by measuring the decrease in mass of a sputtering target. Normally such experiments are performed using ion-beam sputtering, in a fairly high-vacuum environment. The target area bombarded by the beam is small, and fairly low sputtering rates are achieved, but it has the advantage that ion energies are accurately given by the accelerating potential. The methods of sputtering normally used for the preparation of thin films are based on the establishment of a population of positive ions in a low-pressure gaseous environment. Such techniques depend on the establishment of a glow discharge, and require a considerably higher operating pressure, whereas ion-beam sputtering does not require the presence of gas molecules in the deposition chamber, and can therefore be used under UHV conditions, where gaseous contamination effects are minimised. A complicating factor, which often has to be taken into account, is that some incident ions may become embedded in the target, thus tending to increase its mass. The net loss of material from the target Δm is then the difference between the mass of material sputtered and the mass gain of the target. If N_{ion} and N_s are the incident ion impinge-

ment rate and the atomic sputtering rate of the target respectively (29.18), and m_1 and m_2 are the incident ion and target atom masses, it is simple to calculate the mass loss and increase of a given target with bombarded area A sputtered for a time t . The loss of mass is given by $N_s m_2 A t$, and the increase in mass is given by $N_{ion} \Gamma m_1 A t$, where Γ is the probability that an incident ion becomes embedded in the target. Hence

$$\Delta m = N_s m_2 A t - N_{ion} \Gamma m_1 A t \quad (29.29)$$

or

$$S(E) = \frac{N_s}{N_{ion}} = \frac{\Delta m}{m_2 N_{ion} A t} + \Gamma \frac{m_1}{m_2}. \quad (29.30)$$

Frequently the second term on the right-hand side of this equation is negligible, and the sputtering yield is then given by the first term only. This is particularly applicable for large ion doses, while for intermediate doses a detailed empirical knowledge of the value of Γ is required. For low ion doses the approximation $\Gamma = 1 - R_0$, where R_0 is an ion reflection coefficient, may be used [29.53]. It is often useful to express the sputtering yield in terms of the ion current at the cathode $I_c = e N_{ion} A$, where e is the electronic charge and the sputtering yield is then given by [29.1]

$$S(E) \approx \frac{e \Delta m}{m_2 I_c t}. \quad (29.31)$$

Measurements of the ion current, target mass loss and the sputtering time are then all that are necessary to calculate $S(E)$. Although strictly speaking the above expressions are only applicable for sputtering performed with ions of a single energy, they may also be applied to a first approximation for practical sputtering methods using the glow discharge.

Although ion-beam sputtering is a very useful technique for measuring sputtering yields and comparing with theoretical predictions, it is not normally used for routine film deposition, owing to the need for UHV conditions, the low sputtering rate obtained and the associated expense. Virtually all sputtering methods used for the deposition of thin films utilise the glow discharge phenomenon. When a gas at reduced pressure is subjected to an applied voltage, usually in the range of a few hundred to several kilovolts, any free electrons will be accelerated in the field, acquiring energy and ionising the gas molecules to produce a plasma. Under the influence of the field the positively ionised gas molecules are accelerated towards the cathode and free electrons towards the anode. The gas pressure required for the establishment of such a discharge is of the order

of a few pascals, which in turn is determined by the molecular concentration required to maintain the mean free path within suitable limits to initiate and sustain the discharge. The basic principles of the glow discharge are well known and documented [29.55, 56] and need not concern us here. However, it should be noted that the positive ions tend to accumulate in front of the cathode (the cathode-fall region), across which most of the applied voltage is dropped. Thus the electric field between the cathode and the anode is distinctly nonlinear, and most of the energy acquired by the ions in their journey to the cathode is obtained in the cathode-fall region.

It is relatively simple to set up a self-sustained glow discharge, although the behaviour of the discharge depends in a fairly complex manner on the pressure, the applied voltage and the geometry of the system. Inert gases such as argon are generally used, unless the sputtered film relies on a reaction between the sputtering gas and the sputtered material to produce the required compound film. Once the discharge is established, the cathode is continuously bombarded by energetic ions, and provided the energy of these ions is sufficient, sputtering of the cathode will occur. Thus if the target is sufficiently conductive, all that is necessary is to make the cathode of the target material. Substrates on which deposition will occur are normally placed parallel to the target at the anode so that the sputtered atoms need only cover a short distance before deposition. The energy of the sputtered atoms when they reach the substrate is considerably higher than in the case of evaporation, leading to better adhesion. Furthermore, sputtered compounds are normally deposited without dissociation, and therefore the stoichiometry of the target material is usually preserved in the deposited film. This basic form of sputtering is normally termed *diode sputtering*, since only two electrical connections are required.

Basic diode sputtering operates reasonably satisfactorily, but has a major disadvantage in that the sputtering rate N_s depends directly on the impingement rate of the incident ions N_{ion} and thus also on the discharge current. At low gas pressures the supply of ions is limited and thus the discharge current and the sputtering rate are low. Moreover, at sufficiently low pressures the electrons do not ionise a sufficient number of gas atoms to sustain the glow discharge. This low-pressure limit is in fact a major drawback in diode sputtering, since enforced operation at higher pressures above about 3 Pa results in significant film contamination. For this reason sputtering was for many years rejected in favour of evaporation for thin-film deposition. A major improvement in sputtering technology, which enabled sputtering

to be performed at lower pressures was the introduction of *triode sputtering*. In this technique, the discharge is maintained by using an additional source of electrons, so that it is not totally dependent on secondary electron emission from the cathode. Electrons are supplied by an additional thermionic emitter, consisting of a filament through which a current is passed. The electron concentration in the plasma can be readily controlled by varying the anode potential or the filament current, while the ion current is directly controlled by the target potential. A further innovation usually employed in triode sputtering systems is the use of a coil to set up a magnetic field, which forces electrons from the emitter to travel in spiral paths, thus greatly increasing the distance travelled from the emitter before they are collected by the anode. This enhances the ionisation probability α , and thus the supply of positive ions.

It should be emphasised that both diode and triode sputtering may not be used for the sputtering of insulators – a very important class of materials. Since positive ions are responsible for the sputtering process, these need to be neutralised by electrons from the external circuit. In the case of insulators a positive charge builds up at the surface of the target, and this charge cannot leak away through the insulator. Sputtering ceases when the positive charge causes the potential at the target surface to approach that of the plasma. Various methods, such as bombarding the target surface with electrons, have been attempted to solve this problem, but in general with little success. Glow discharge sputtering of insulators is generally therefore performed reactively, or with alternating radio-frequency (RF) fields, as described later.

As with evaporated films, there are a number of problems associated with the deposition of alloys and compounds. The best theories of sputtering yields are typically inaccurate by a factor of two or so. Empirical determinations are normally required for films of controlled composition to be deposited. Nevertheless, sputtering of multicomponent or compound targets is frequently all that is necessary to obtain films varying only marginally from the target composition. *Maisel* [29.57] suggests that during the initial sputtering run an *altered region* is formed on the surface of the target, which is deficient in the target component having the highest sputtering yield. Thus although stoichiometric films will not normally be deposited during the initial sputtering run, such films may generally be deposited during subsequent runs, since the deficiency in the higher sputtering rate material compensates for its greater removal rate [29.51, 57]. Minor differences between the compositions of the film and target do still

occur, but are normally the result of oxidation or of the evaporation of one component at high temperatures. When compounds are sputtered the incident ion energies are often sufficient to break the chemical bonds, again leading to a deficiency in one component. *Reactive sputtering* and *co-sputtering* are specifically designed to give greater flexibility during the deposition of multicomponent films.

Reactive sputtering involves the sputtering of a target using a sputtering gas which is amenable to chemical combination with the target material. The gas used may be either solely the reactive species, or a mixture of the reactive species and an inert sputtering gas, such as argon. Clearly the process is suitable for preparation of certain insulators which cannot be prepared by the previously mentioned sputtering techniques. Deposition rate depends primarily on the target sputtering yield, through the energy of the incident sputtering ions, the gas pressure and the distance between the target and the substrate. For example, *Pernay et al.* [29.58] have concluded that the deposition rate is primarily determined by a reduced electric field E^* , given by E/p , where E is the mean electric field between the cathode and the anode, and p is the gas pressure. When a copper target was sputtered in a mixture of oxygen and argon, either Cu_2O , CuO or Cu were deposited depending on the value of E^* and the oxygen concentration [29.59]. Reactive sputtering has been used mainly to deposit oxide films which have insulating or semiconducting properties. Among the common oxides that have been studied are those of aluminium, cadmium, copper, iron, silicon, tantalum, tin and titanium. Oxides of niobium, thorium, vanadium and zirconium have also been investigated, as have the rare-earth oxides of hafnium, lanthanum and yttrium. The other major non-oxide compounds which have been investigated are the nitrides, with silicon nitride, Si_3N_4 , receiving most attention due to its utility both as an insulator and in variants of the silicon planar process for integrated-circuit fabrication. This material has been prepared using a Si_3N_4 target and a mixture of argon and nitrogen as the sputtering gas [29.60]. Carbides and sulphides have also been sputtered reactively. In all cases it is imperative that the stoichiometry of the deposited films is determined using the normal chemical techniques, and the deposition parameters adjusted until the required composition is attained. A very useful review of reactive sputtering, covering both the mechanisms and the primary reactive sputtering techniques, has been given by *Westwood* [29.61].

Co-sputtering has certain advantages from the point of view of the preparation of films whose components

have different sputtering yields. Segregation of components on cooling can lead to a nonstoichiometric composition of material in the target, and differences in sticking coefficients between components can lead to the deposited composition differing from that arriving at the substrate. Co-sputtering has the advantage that the proportion of each sputtered components may be controlled during deposition, and may readily be adjusted in response to varying conditions to ensure stoichiometry of the deposited film. Co-sputtering may be performed with two or more conventional sputtering targets operating simultaneously for each component, each directed towards a common substrate. Another method of co-sputtering is again to use conventional targets, but with the substrate alternately subjected to the sputtered material from each target. This may be achieved by mounting the substrate on a disc which rotates in front of the targets. Alternatively a specially designed target assembly may be used such that the components arrive at the substrate in their stoichiometric proportions; the potentials of the components may be independently controlled to compensate for differing sputtering yields and sticking coefficients. Generally a plane-parallel substrate arrangement, with the target large in comparison to the target-substrate distance is employed, which maximises the uniformity of the film thickness. *Sinclair and Peters* [29.62] have described some target-electrode assemblies which have proven suitable for co-sputtering, including a concentric ring and disc assembly and an interdigitated design. Each of these allows independent control of the voltage applied to the two parts of the electrode assemblies and thus control over the composition of the sputtered films. Although the spatial dependence of the electric field in such assemblies is complex, and the deposition rate of one component may be influenced by the potential applied to the electrode for the other component, it is relatively straightforward to calibrate such a source for different materials, and to compensate for such effects by varying the acceleration potentials accordingly.

It is clear that, in general, impurities present in the gaseous phase may be incorporated into the deposited films. The fraction f_i of the impurity species i trapped within a film is given by [29.57]

$$f_i = \frac{s_i N_i}{s_i N_i + N_R} \quad (29.32)$$

where s_i and N_i ($\text{m}^{-2} \text{s}^{-1}$) are respectively the sticking coefficient and impingement rate of the impurity species and N_R is the molecular deposition rate of the required film material; f_i can be reduced by increas-

ing N_R , or by decreasing either or both of s_i or N_i . In *bias sputtering* the phenomenon of re-sputtering is employed, i. e. components of the sputtered film are re-sputtered by ions from the plasma. This is achieved by applying an additional small negative potential to the substrate. Preferential sputtering of the impurities takes place, providing the binding energy of the impurities is lower than that of the host film material. This leads to an effective reduction in f_i , below that predicted by (29.32) [29.63]. The value of s_i in the expression is effectively reduced by re-sputtering of impurities from the film surface. *Asymmetric alternating current (AC) sputtering* [29.64] achieves similar objectives to bias sputtering, i. e. bombardment of the cathode target by ions to cause sputtering and bombardment of the films by lower-energy ions to preferentially re-sputter the impurities. These are accomplished by establishing an alternating current between the substrate and cathode and the use of a simple diode-resistance network. During the negative half-cycle the full negative potential is applied to the cathode, allowing sputtering to take place. In the following positive half-cycle, a relatively small positive potential (limited by the resistance) is applied to the cathode. This allows low-voltage sputtering of the films, which is sufficient to remove only the gaseous impurities. In this technique, again the value of s_i is effectively reduced, leading to a decrease in f_i . In *getter sputtering*, the gettering action of a freshly deposited thin film is utilised to purge the chamber of unwanted active gases. The concentration of impurities in thin films has been observed to decrease as the film thickness increases, and is attributed to the trapping (or gettering) action of the film material. In getter sputtering systems this action is exploited so that the majority of impurities are trapped in a film region deposited between the gas entry point and the substrates, while the required film is deposited elsewhere in the chamber where the gas impurity concentration will already have been depleted by the gettering action. The design of a typical getter sputtering system is described by Theuerer and Hauser [29.65]. In the case of getter sputtering, N_i is effectively reduced in (29.32), lowering the value of f_i .

Trapping of inert gas molecules in the films is also a possibility, and their concentration should be given by (29.32). However, at thermal energies the sticking coefficient is effectively zero, and thus trapping ought not to be of great concern. Nevertheless, for sputtering systems using argon, concentrations of 0.1% have been observed [29.67], generally for ion energies greater than 100 eV. Although not a major problem for many appli-

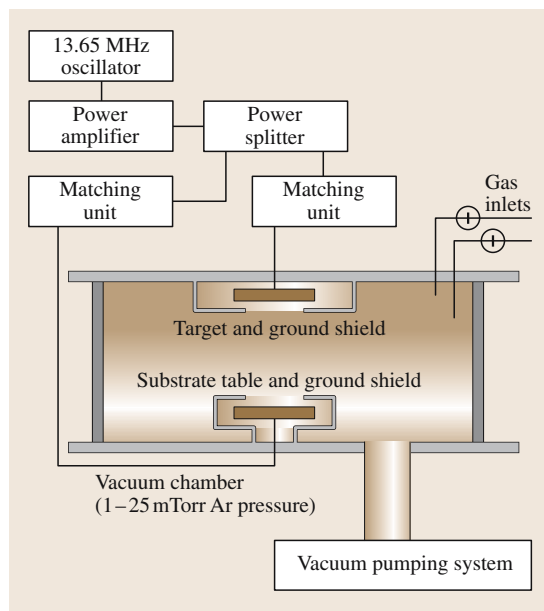


Fig. 29.6 A typical radio-frequency (RF) sputtering system. After [29.66] with permission from Oxford Univ. Press

cations the possibility of inert-gas contamination should not therefore be discounted.

It has been mentioned earlier that simple direct current (DC) discharges as used in the diode and triode sputtering techniques described above, are unable to deposit insulating films. Although reactive sputtering may be used, the deposition of films by this method requires careful control of sputtering rates and gas pressure, and often calibrations of individual systems and materials. The use of *radio-frequency (RF) sputtering*, allows insulators to be sputtered directly [29.68]. When a high-frequency (typically several MHz) alternating signal is applied to a metal plate, on which an insulating (dielectric) target is mounted and is placed within an auxiliary low-pressure discharge, the positive charge which accumulates on the target is neutralised during half of each cycle. Owing to the much higher mobility of the electrons than the ions in the plasma, the plasma potential becomes positive with respect to the target. RF sputtering was first employed by Davidse and Maisel [29.69], who found that a self-sustained discharge could be started and maintained by the application of RF power to the electrode only. The mechanism by which sputtering takes place requires the frequency to be in the radio-frequency range [29.16]. The electron mobility is several orders of magnitude greater than that of the ions

which have difficulty in keeping up with the field variations, and at high frequencies the ions accumulate in front of the target, in the same way that they do in a DC discharge. Because the number of electrons arriving at the target during the positive half-cycle are not matched by the number of ions arriving during the negative half cycle, the front of the target acquires a negative charge, which repels most of the electrons that would arrive at the surface during the positive half-cycle. The plasma itself is said to be positively *self-biased* with respect to the target. It is therefore possible to both feed energy into the plasma, while also giving the ions sufficient energy to cause sputtering. The RF discharge is self-sustained, and Davidse and Maissel found that an externally generated plasma is not required. A schematic diagram of the RF sputtering equipment is shown in Fig. 29.6 [29.66] while more detailed discussion of the RF sputtering process is given in previous reviews [29.16, 57]. These authors point out that a major difference between diode and RF sputtering is a requirement for an impedance matching network or *match box* [29.70] in the RF case. The matching network is inserted between the power supply and the discharge chamber, its purpose being to maximise the power that can be delivered to the plasma. Typically a network containing an inductor and a capacitor is used. An example is shown in the literature [29.71]. A blocking capacitor is normally inserted in the network to ensure that the build up of the self-bias voltage proceeds satisfactorily. Although the target itself will normally perform this function in the case of insulator sputtering, the inclusion of a blocking capacitor allows RF sputtering to be used for the deposition of conductors as well as insulators, so semiconductors and metals are also easily deposited by this technique. Mainly for this reason RF sputtering has become a versatile technique which is used near-universally for semiconductor production as well as for insulators and metals [29.70]. A significant increase in the sputtering rate is also observed in RF systems.

Finally, a technique used to increase the sputtering rate still further is that of *magnetron sputtering*. In this technique, a magnetic field is applied between the target and the substrate. Electrons in the plasma then experience the Lorentz force from the magnetic field in addition to that derived from the electric field, so that the total force on the electron is given by

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = e\mathbf{E} + e\mathbf{v} \times \mathbf{B}, \quad (29.33)$$

where \mathbf{F} is the total force experienced by the electron, m and e are the electron mass and charge respectively,

\mathbf{v} is the electron velocity, \mathbf{E} is the electric field and \mathbf{B} is the magnetic field [29.19, 72]. The effect of the two fields is to cause the electrons to move in spiral rather than linear paths, and therefore to greatly increase the distance travelled by the electrons in the plasma. This substantially increases the time the electrons are within the plasma, and thus the probability of their causing secondary ionisation is also significantly enhanced. In turn this leads to increased current flow and crucially increased deposition rates. The magnetron sputtering source is constructed by mounting a powerful magnet below the target, with one pole concentric with the other as shown in Fig. 29.7 [29.66]. This establishes a magnetic field across the target surface; the direction of the magnetic field is perpendicular to the electric field at the furthest extent of the magnetic force lines from the target. The result of the two interacting fields is to confine the electrons for a significant period within a region termed the *racetrack* at the surface of the target, where the electrons effectively hop across the surface. Ionisation of the sputtering gas occurs mainly above the racetrack, and erosion of the target takes place preferentially in this region. In addition to the enhanced deposition rate, magnetron sputtering also has several other advantages. Since the ionisation region is effec-

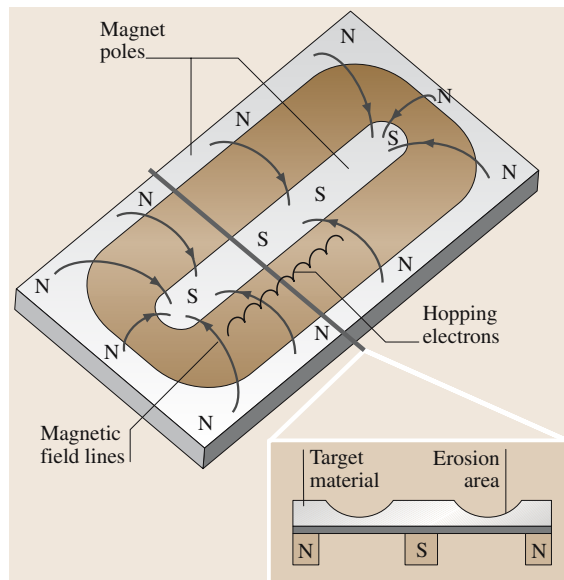


Fig. 29.7 The technique of magnetron sputtering, showing the hopping electrons of the target confined to a *racetrack* region by the magnetic fields from the magnets under the target. After [29.66] with permission from Oxford Univ. Press

tively confined to the racetrack, unwanted sputtering from the substrate and the walls of the chamber are reduced, as is the substrate temperature. The gas pressure required to sustain the discharge is also reduced, owing to the greater ionisation efficiency.

A further improvement in the technique of magnetron sputtering may be obtained by the use of an unbalanced magnetron, as described by *Window and Savvides* [29.73]. In spite of its excellent high deposition rate, magnetron sputtering generally produces films which have a columnar structure. This leads to the presence of voids, both within grains and at grain boundaries. Low-energy ion bombardment has been shown to improve adatom mobility, allowing atoms to migrate on the depositing surface and filling some of the voids [29.72]. Various computer simulations described in this reference have illustrated the importance of ion beams in depositing dense, void-free films. It is possible to generate an ion current in the plasma by using an unbalanced magnetron where the inner and outer magnets are deliberately unbalanced from the ideal magnetron configuration. Further details concerning the field characteristics of unbalanced magnetrons and of the unbalanced magnetron technique are given in the literature [29.72, 73]. The method has been used to successfully deposit various types of film, including those intended for electrical and optical applications, corrosion protection, and wear and abrasion resistance. A fuller discussion of materials prepared using this method is given by *Rohde* [29.72].

Magnetron sputtering may also be used in conjunction with other sputtering techniques. A combination of RF and magnetron sputtering is frequently used, for example for Co:M alloys where M represents Al, Si, Ti, Cr, Fe, Ni, Cu, Zr, Nb, Mo, Ag, Ta or W [29.74], and also for the deposition of silicon nitride films [29.60].

Thus sputtering has developed from a relatively simple DC technique capable only of the deposition of conductors at a relatively low rate, to encompass both conductors and insulators (RF sputtering) and with sputtering rates suitable for high-throughput industrial processes (magnetron sputtering). Improvements in film morphology have been accomplished by the use of unbalanced magnetron sputtering.

29.1.2 Chemical Deposition Methods

There are many different chemical methods for preparing thin films, but by their nature many of these are restricted to one or perhaps a few different materials, related to a specific chemical reaction or series of reactions. It is therefore only possible to categorise some

general techniques and to give examples for various particular materials. *Campbell* [29.75] divided chemical methods into two classes: those depending on an electrical source of ions, such as electroplating and anodisation, and those requiring a chemical reaction, usually, but not exclusively involving a chemical vapour as in the CVD method. A further technique, which has found favour in recent years, is the deposition of molecular films using the Langmuir–Blodgett technique. Brief examples of these are given below.

Electrodeposition

By electrodeposition, we include all methods which involve passing an electric current through a solution such that a film of a material, usually a metal or an oxide, is deposited on one of the electrodes immersed in the solution. Electrolytic deposition, or *electroplating*, is the deposition of a metal by this method. According to *Campbell* [29.75, 76], of the 70 metallic elements, 33 can be electroplated successfully, although of these only 14 are deposited commercially. The metallic film is deposited on the cathode, and its thickness and the time in which this occurs are determined by the two laws of electrolysis:

1. The mass of the deposit is proportional to the quantity of electricity passed.
2. The mass of material deposited by the same quantity of electricity is proportional to the electrochemical equivalent, E .

This can be expressed as

$$\frac{W}{A} = JtE\alpha, \quad (29.34)$$

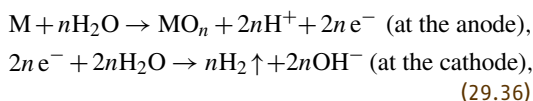
where W is the mass deposited on a cathode of surface area A , J is the current density, t is the time of deposition and α is the *current efficiency*, or ratio of the expected to the theoretical mass deposited. Its value is generally in the range 0.5–1.0. The previous expression can be rewritten in terms of the film deposition rate, R (or thickness d per unit time t), by noting that the film density $\rho = W/Ad$. Combining this with (29.34) gives

$$R = \frac{d}{t} = \frac{JE\alpha}{\rho}. \quad (29.35)$$

When a high current is passed through a suitable electrolyte, metallic ions move towards the cathode under the influence of the applied electric field. The mass deposited, and the deposition rate are given by (29.34) and (29.35), and may easily be estimated if E and ρ are known. These are tabulated for the 14 metals commonly

deposited by this method [29.75]. Deposition rates can be very high, depending on the current drawn through the electrolyte. For example, silver will deposit at a rate of about 1 nm/s for a current density of 10 A/m², rising to 1 μm/s at 10 000 A/m² [29.75]. Further details of the deposition of alloys are given in a standard text [29.77], while the growth and structures of electrodeposits are described by *Lawless* [29.78].

A second technique which is used for the deposition of oxides (or less commonly for other compounds such as nitrides), is that of *anodisation*. As its name implies, films are deposited on the anode of the parent metal, following a chemical reaction. The chemical equations describing the anodisation process are given below [29.75]:



where M represents the metal and n is an integer specifying the oxide deposited. The oxide grows in an amorphous form at the metallic anode surface, while hydrogen is evolved at the cathode. The electrolyte is typically an aqueous solution of water containing a dilute acid.

In general a constant current is passed through the cell and film thickness is proportional to the time elapsed. Films grown under constant-voltage conditions, show a decreasing deposition rate with time, as the current falls to zero when the voltage dropped across the film increases to the total applied voltage. Thus the maximum thickness of the film is determined by the voltage used. Growth rates for anodic films under constant-current conditions are of the same order as those for the *PVD* methods of evaporation and sputtering. Aluminium and tantalum anodisation at a current density of 20 A/m² gives growth rates of the order of 1 nm/s [29.76]. Considerably more detailed discussions of anodic oxide films in particular are given in the literature [29.79, 80]. These cover considerably more details of the types of films that may be successfully prepared, their structures, characterisation methods and copious references to related work.

Chemical Vapour Deposition (CVD)

Chemical vapour deposition is a method whereby a volatile compound reacts, with other gas species, to produce the required compound on the substrate. *CVD* does not require operation under vacuum, as with the *PVD* processes of evaporation, *MBE* and sputter-

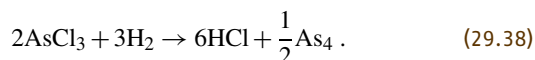
ing. Moreover, it is capable of producing both silicon and gallium arsenide (GaAs) semiconductor films, and to grow these on substrates of the same material as high-quality epitaxial layers (homoeptaxy). For these reasons it has found wide acceptance in semiconductor processing. High deposition rates relative to *PVD* deposition methods are possible, and controlled doping performed during the deposition process is relatively straightforward. Objects having quite complex shapes can be coated relatively easily, since the film deposition does not depend on line-of-sight geometry as in the case of evaporation and to a lesser extent sputtering. There are however some formidable disadvantages with the technique. Generally much higher substrate temperatures are required in *CVD* than for the other methods. The reactive gases used in the deposition process and/or the reaction products can be extremely toxic, explosive or corrosive, requiring considerable investment in safety features. This may well be justified in high-throughput semiconductor production processes, but uneconomic for smaller-scale development work, where *PVD* methods are preferable. Corrosion and effects of unwanted diffusion, alloying and chemical reactions at the substrate surface may occur under the high operating temperatures, and masking of substrates is particularly difficult [29.81].

In principle *CVD* may utilise any chemical reaction which produces the required material, and therefore the equipment used for the production of a given material is likely to differ from that required for another material. Nevertheless, there are several different classes of reaction which are used in *CVD* processes which are described in the literature [29.16, 18, 19, 75, 76]. Some of these are outlined below.

Pyrolysis or thermal decomposition of a gaseous compound entails passing a vapour over a heated substrate, which causes decomposition and the condensation of a stable solid. This has been used primarily in the production of silicon from SiH₄ and nickel from nickel carbonyl. *Oxidation* is frequently employed to provide SiO₂ in silicon processing. Halides of the required metal oxide are reacted with steam to give the oxide and a hydrogen halide. In *nitriding* the steam is replaced with ammonia (NH₃), and nitrides such as Si₃N₄ may be generated. *Reduction* reactions use hydrogen gas instead of steam or ammonia in the preparation of silicon and refractive metals such as tantalum or molybdenum from their halides. In *disproportionation*, a reversible reaction of the type



is usually employed [29.76], where A represents a metal, B is usually a halide and both AB and AB_2 are gases. The higher valence state is more stable at lower temperatures so, if the hot gas AB is passed into a colder region, AB_2 will be formed and A deposited. In *transfer reactions* for the preparation of compounds, two different strategies may be employed. Either a compound can be used as the source material, or the source comprises a volatile compound of one of the elements of the required compound. Gallium arsenide (GaAs) can be prepared by both methods [29.16, 83]. In Fig. 29.8 [29.82] a typical CVD system is shown which uses a transfer reactor of the second type to produce GaAs. Highly purified arsenous chloride ($AsCl_3$) and Ga are the source materials. Purified hydrogen is bubbled through the $AsCl_3$ and passed through the furnace. In zone 1 of the furnace, the following reaction takes place:



The arsenic vapour is completely absorbed by the Ga source in zone 2 until saturation occurs at 2.25 at. % As. No free arsenic condenses beyond the furnace, but gallium is transported as a lower chloride. Deposition of GaAs occurs in zone 3, when hydrogen is diverted to bypass the bubbler.

Tables of materials which can be deposited by CVD and the associated reactions are given in the literature [29.19, 75]. However, several modifications to the basic CVD process are also in current use. These seek to overcome some of the deficiencies associated with CVD, such as the operation at atmospheric pressure and the use of very high temperatures often involved. Systems employing organic precursors also have certain advantages for the deposition of various materials.

We first discuss low-pressure CVD or LPCVD. Typical pressures of the order of 100 Pa are used, approximately a thousandfold reduction from atmospheric pressure. Contamination is normally lower than for atmospheric CVD and film stoichiometry is often improved. To compensate for the reduced pressure, the input reactant gas concentrations are increased, the gas flow velocity is increased, but the gas density is reduced. Because the mean free path of the gas molecules increases due to the reduced pressure, substrates can be placed closer together, giving a higher throughput in commercial applications. Silicon and dielectric films are generally prepared by this method. The operating temperatures required for LPCVD remain high, however.

In plasma-enhanced CVD (PECVD) a glow discharge is initiated inside the chamber where the reaction occurs. A comprehensive review of the PECVD technique is given by Ohja [29.84]. Dissociation of the gas in the plasma takes place, owing to the nonequilibrium nature of the plasma, which has a high electron temperature but a low gas temperature. The main advantage of this technique is that the plasma decomposition allows film deposition to take place at much lower temperatures than those employed in atmospheric CVD using decomposition reactions and also in LPCVD. In PECVD, as in LPCVD, the pressure is reduced considerably below atmospheric pressure, although the range of pressures employed varies over a wider range, and can be as high as a few hundred Pa. The discharge is usually excited by an RF field, although both DC and microwave frequency fields have also been used. Average electron energies are of the order of several eV, which is sufficient to cause ionisation and decomposition of the gas molecules. A major application of PECVD is in the deposition of silicon nitride films, an important dielectric used in semiconductor processing for microelectronic

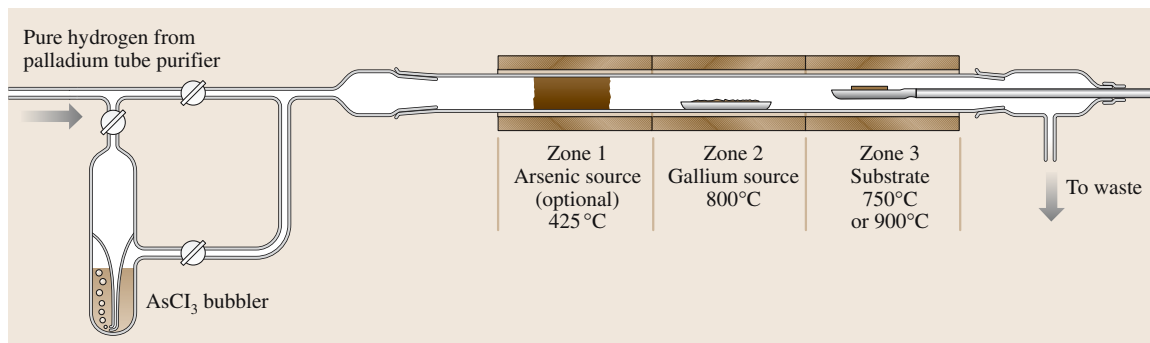


Fig. 29.8 Epitaxial growth apparatus for the production of GaAs using chemical vapour deposition (CVD). After [29.82] with permission from Elsevier Science

passivation. Here the temperatures may not exceed about 300 °C, otherwise damage to the circuitry can result, and so atmospheric CVD and indeed LPCVD are not suitable techniques. For details of the chemical reactions involved, and a comparison of the properties of silicon nitride prepared by LPCVD and by PECVD the reader is referred to the literature [29.19, 85]. The former reference also includes a table giving details for the deposition of many other thin-film materials, including silicon, germanium and other semiconductors, as well as various oxides, nitrides and carbides. PECVD systems operating at microwave frequencies utilise the phenomenon of electron cyclotron resonance (ECR); this deposition method is known as electron cyclotron PECVD (or ECR-PECVD). Frequencies are of the order of 1–10 GHz and the ECR plasma may be generated at pressures of about 10^{-3} – 10^{-1} Pa, leading to a degree of ionisation up to about 1000 times that achieved in an RF plasma [29.19]. High deposition rates, low-pressure operation and absence of contamination have made ECRPECVD a very attractive technique for film deposition. A full review of this technique has been given by Popov [29.86].

Finally we briefly mention the technique of metalorganic CVD (MOCVD),

which is becoming an increasingly important method of film deposition. The technique is essentially the same as CVD, except that at least one of the precursors is a volatile metalorganic compound, such as a metal alkyl. The formation of the required compound takes place by pyrolysis of the metalorganic compound with another (inorganic or organic) gaseous precursor. The main advantage of MOCVD is the relatively high volatility of the organic compounds at moderate temperatures, and the fact that, since all these are in the vapour phase, precise control of gas flow rates and partial pressures are possible [29.19]. The compound semiconductors such as GaAs, GaP and InP may be prepared using, for example, trimethyl-Ga [(CH₃)₃Ga] or trimethyl-In [(CH₃)₃In] as precursors. The basic reaction is with a Group V hydride, e.g. AsH₃ or PH₃ for the compounds listed above. A short table of some film materials which have been successfully generated by MOCVD is given in the literature [29.19]. In addition to its applications in preparing compound semiconductor films MOCVD has also been applied to the deposition of high-temperature superconducting thin films, with quite complex structures [29.88].

Langmuir–Blodgett Film Deposition

The Langmuir–Blodgett technique of film deposition entails the coating of solid substrates with molecular

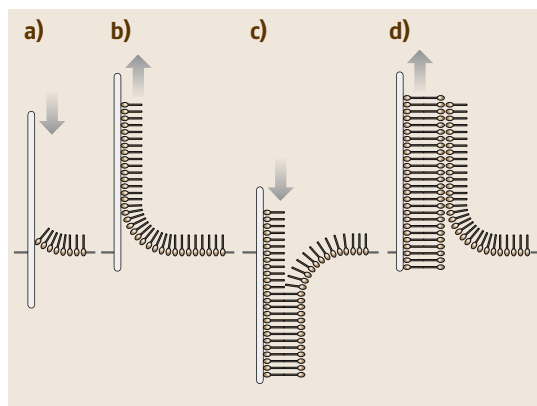


Fig. 29.9a–d Langmuir film deposition: (a) immersion of a hydrophilic substrate without film deposition; (b) withdrawal of the hydrophilic substrate; (c) repeated immersion with film deposition; (d) continuation of the process. After [29.87] with permission from Elsevier Science

layers. Langmuir [29.89] first described the transfer of fatty-acid organic monolayers from the surface of water to a rigid substrate, while Blodgett [29.90] extended the work to show that the process could be repeated so that monolayers were sequentially transferred to form multilayer films. The early history of Langmuir–Blodgett films has been briefly described in the preface to the Proceedings of the First International Conference on Langmuir–Blodgett Films [29.91]. In the basic deposition process, a fatty-acid film is spread over the surface of a subphase (usually purified water). Most molecular-film materials have a molecular structure where one end of the molecule is hydrophilic, which seeks out the water subphase, while the other end of the molecule is hydrophobic and attempts to point away from the water subphase. We therefore have the situation where the hydrophilic end of the molecule is oriented downwards towards the water, with the hydrophobic end oriented upwards. In Langmuir–Blodgett deposition, the required molecules are dispersed in a volatile solution, which eventually evaporates, leaving the molecules with their hydrophilic ends immersed in the water, but with their axes randomly oriented [29.87]. A compressive surface-barrier system is employed to compress the molecules so that they are close-packed and oriented perpendicular to the surface of the subphase, with their hydrophilic ends pointing vertically downwards. The molecules can then be transferred to either a hydrophilic or hydrophobic substrate by slowly dipping into the subphase. The basis process is shown in Fig. 29.9 [29.87], for the case of a hydrophilic substrate. On the initial introduction of

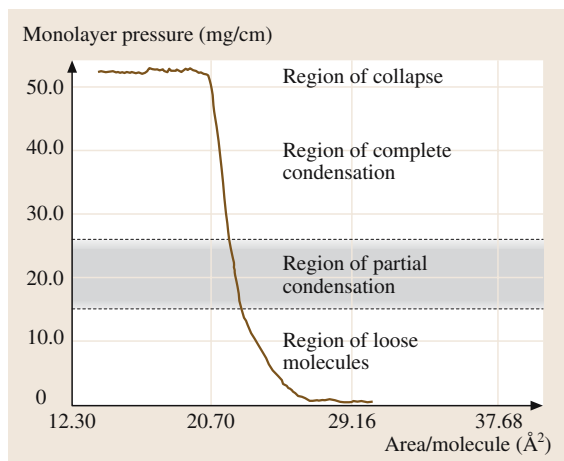


Fig. 29.10 The pressure–area relationship for a cadmium stearate molecular layer obtained at a temperature of 22 °C at pH 5.6. Film deposition should be carried out in the linear region of the curve, i. e. the region of complete condensation without collapse. After [29.87] with permission from Elsevier Science

the substrate into the subphase no coating takes place (a), but on removing it a film attaches to the substrate by polar-bond formation (b). The substrate is then effective hydrophobic, and on dipping for a second time, the hydrophobic ends of the molecules on the subphase form van der Waals bonds with the hydrophobic substrate surface, thus again making the surface hydrophilic (c). On withdrawing the substrate for a second time a further molecular layer is deposited, with the hydrophobic ends of the molecules at the surface (d). Repeated dipping cycles then establish multilayer films with an odd number of layers. The use of an initially hydrophobic substrate gives an even number of molecular layers for each completed dipping cycle.

It is evident that we have a mechanism whereby a predicted number of molecular layers may be deposited, and knowing the molecule dimensions (i. e. the length of the molecule between its hydrophilic and hydrophobic ends) together with the number of dipping cycles, it is easy to accurately calculate the film thickness. It is important, however, that the surface pressure is maintained within a fixed range of values during deposition. If it is too high the surface film collapses in localised regions, giving a transferred layer thickness greater than one monolayer; conversely, if the pressure is too low, there may be voids in the film surface, which transfer on to the substrate. This is illustrated in Fig. 29.10 [29.87], which shows a pressure–area isotherm for a cadmium

stearate molecular layer. It is normal to use a feedback system during deposition to control a moveable barrier on the film surface to maintain the surface pressure within a narrow range while the dipping process takes place. Here the molecular area remains effectively constant for a large change in pressure (region of complete condensation in the figure), ensuring that small changes in pressure do not cause the deposited film to fold over or break up.

There are, however, some limitations to the use of the Langmuir–Blodgett deposition method. The material must satisfy the following requirements [29.87]. The molecular chain structure needs to be hydrophilic at one end and hydrophobic at the other. Furthermore, it should be insoluble in water, but soluble in a suitable solvent in order to give a spreading solution. Some materials which are insoluble can, however, be modified by chemical substitution of molecular end groups. For example, metal-free phthalocyanine is normally insoluble, but the chemical addition of $C(CH_3)_3$ end groups to the molecule renders it soluble in some aromatic solutions such as toluene. This type of chemical modification allows the deposition of materials which cannot normally be deposited to be deposited in a substituted form [29.92]. Many other substituted organic compounds may also be deposited, including dyes, porphyrins, fullerenes, charge-transfer complexes and polymers [29.93]. It is also possible to prepare Langmuir–Blodgett film layers with alternating monolayer structures, by dividing the Langmuir–Blodgett trough into two compartments, one for each type of monolayer, separated by a surface barrier [29.94,95]. Deposition then takes place by alternately removing the substrate from the surfaces of the two compartments, by passing the substrate through the subphase under the surface barrier between deposition of each of the materials.

The advantages of the Langmuir–Blodgett technique over other thin-film deposition methods lie primarily in the low temperature at which deposition takes place (close to room temperature), and on the feasibility of depositing very thin mono-molecular layers of known thickness. Several types of electronic devices incorporating Langmuir–Blodgett films have been discussed [29.96] as has the writing of very fine lines of width less than 10 nm separated by 20 nm in manganese stearate films [29.97]. Further work on the deposition of other materials by this method may well lead to novel device structures and to improved processing methods. An excellent discussion on the preparation and major properties of Langmuir–Blodgett films is given by Petty [29.93].

29.2 Structure

In this section some examples of the structures of various types of thin films are described. By the *structure*, we mean mainly the crystallographic form in which the film is deposited, and also factors associated with this, such as the mean size of crystallites and their orientations. However, the gross morphology of films is also included, and this covers features generally larger than the grain size. The basic type of film, i. e. epitaxial, polycrystalline or amorphous, is usually deduced by diffraction methods (X-ray, electron or neutron), although we will mainly be concerned with the first of these only. Thus a brief description of the method of X-ray crystallography is first given, followed by examples of different types of films that have been observed. Finally some examples of the larger-scale morphology are described.

Several reviews of film structure have been given previously [29.98–100], as well as more generalised discussions in most of the texts covering thin films, to which the reader is referred. However, it should be noted that the large number of variables encountered in thin-film deposition processes (e.g. substrate temperature, deposition rate, angle between vapour stream and substrate) permit films of differing structures to be deposited under nominally similar conditions. It is always wise, therefore, before interpreting measurements of any thin-film properties (such as the electrical or optical properties) to first fully determine the structure of the films, and to take this into account in any interpretation.

29.2.1 Crystallography

In general films tend to deposit from the vapour phase in one or more of their documented bulk-crystal structures. In crystals, the spatial arrangement of the atoms exhibits a symmetry, which can be replicated throughout three-dimensional space by repetition of a basic unit, called the primitive unit cell. The unit cell of any crystal structure may be specified by the magnitudes of three vectors **a**, **b** and **c** representing the unit cell lengths, and the angles α , β and γ between the three vectors. Symmetry considerations allow all crystals to be classified into seven crystallographic structures, namely the cubic, tetragonal, orthorhombic, hexagonal, trigonal, monoclinic and triclinic forms, of which only the first three are orthogonal. Conditions on the equality or otherwise of the lengths *a*, *b* and *c* and the angles between them determine into which type the structure of a given material falls. In three dimensions the atoms arrange themselves in planes, with gaps between the planes known as the

planar spacing, d_{hkl} . In any perfect crystal there are many sets of crystal planes that may be identified. The possible planar spacings are given by different expressions, depending on the crystal structures. For example, in the cubic structure ($a = b = c$, $\alpha = \beta = \gamma = 90^\circ$)

$$\frac{1}{d_{hkl}^2} = \frac{h^2 + k^2 + l^2}{a^2} . \quad (29.39)$$

In the tetragonal structure ($a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$)

$$\frac{1}{d_{hkl}^2} = \frac{h^2 + k^2}{a^2} + \frac{l^2}{c^2} \quad (29.40)$$

and in the hexagonal structure ($a = b \neq c$, $c \perp a$, $\gamma = 120^\circ$)

$$\frac{1}{d_{hkl}^2} = \frac{4}{3} \left(\frac{h^2 + hk + k^2}{a^2} \right) + \frac{l^2}{c^2} . \quad (29.41)$$

In these equations *a*, *b* and *c* represent the lengths of the sides of the primitive unit cell, and *h*, *k* and *l* are integers. Similar, but more complex, equations exist for the other crystal structures which have less symmetry. For bulk crystals having unit cell sides of the order of nanometres, the crystal planes act as three-dimensional diffraction gratings, which reflect an incident beam of radiation according to Bragg's law,

$$\lambda = 2d_{hkl} \sin \theta , \quad (29.42)$$

where λ is the wavelength of the radiation, and θ is the angle between the incident beam and the reflecting planes. The values of λ for X-rays, electrons and neutrons may be chosen such that they are of the same order as d_{hkl} , and therefore diffraction will occur at suitable angles. These angles are those which satisfy Bragg's law, and are therefore determined by the radiation wavelength λ and by the integer values (*h*, *k* and *l*) applying to the sets of reflecting planes present. (*hkl*) then specifies the set of reflecting planes, and also the orientation.

For perfect crystals, reflections only occur at certain well-defined angles. If the crystal is rotated with respect to the incident beam of radiation, the deviation angle between the incident and reflected beam is 2θ . If the material, instead of being perfect, is made up of a large number of differently oriented microcrystallites, i. e. it is polycrystalline, the diffraction method will pick up reflections from a large number of individual crystallites, and therefore provide data on all the reflecting planes present. A particular set of planes will only reflect at

a specific value of 2θ , and by observing the reflecting angles the planar spacings d_{hkl} can be determined. By reference to standard tabulated data (such as the *International Tables of Crystallography*), which give details of the reflections and intensities expected for most crystalline materials, a considerable amount of information can be determined. This includes the nature of the phases present, their preferred orientations and a measure of the microcrystallite size.

X-ray diffraction is the most commonly used technique. The penetration depth of X-rays can be quite significant, and if the film is too thin they can penetrate down to the substrate. Thus X-ray diffraction often gives information on the substrate as well as the film. If this is to be avoided, the film thickness needs to be increased or the X-rays directed at a low glancing angle with respect to the substrate surface. Electrons have a considerably lower penetration ability, and therefore give information only on the immediate surface region. Neutron diffraction is used particularly in the investigation of magnetic films.

29.2.2 Film Structure

Apart from the detailed type of crystal structures observed, it is useful to divide the structures of thin films into three main categories, according to their crystallographic perfection (or lack of it). Epitaxial films are essentially perfect films grown on a substrate surface with all the atoms arranged predictably in their correct crystallographic positions. There may be certain defects, such as vacancies, interstitial atoms or impurities, but the deviation from perfection should be relatively small. Furthermore, there may be some slight deviations in the unit cell dimensions from those of the corresponding bulk material, owing to idiosyncrasies of the deposition method or the substrate surface, and in particular to stresses built up in the film during deposition. Various methods have been devised to deposit epitaxial films, such as molecular-beam epitaxy (MBE) and vapour-phase epitaxy, both of which are used in the semiconductor industry for preparation of high-quality films. Non-epitaxial films are frequently polycrystalline, having a large number of grains, oriented in different directions, and separated by grain boundaries. The grain size is heavily influenced by the deposition conditions, and can have a profound effect on secondary properties, such as the electrical conductivity, in which grain boundaries influence the electron mean free path. Finally, amorphous films have no long-range order, being noncrystalline, with only short-range order apparent.

Substrate temperature during deposition and subsequent thermal history after deposition can have a profound effect in determining the initial structure, orientation and subsequent phase changes. Some examples of these effects are included in the following discussion.

In the case of epitaxial films, a large number of factors determine whether the film is indeed epitaxial, and if so the crystalline orientation that will result. The main concepts that underlay the growth of epitaxial films are the equilibrium thermodynamics of the growing nuclei, the evaporation process as described by the Hertz–Knudsen expression (29.5), surface kinetics and mobility on the substrate and the growth of defects [29.101]. The interdependence and interaction of these various processes, which are generally all temperature-dependent, lead in many cases to a particular value of temperature (usually called the epitaxial temperature) only above which epitaxy can take place. Depending on the relative difference between the unit cell dimensions of the substrate and the films (often quantified as the *misfit*) there is frequently a considerable stress between them, which is released by a network of imperfections (dislocations). Details of the epitaxial growth are strongly related to various theories of nucleation. There are several of these, based on the characteristics of microscopic nuclei on the substrate. These grow during the deposition process, and if the conditions are suitable, can result in the deposition of epitaxial films. In the capillarity model [29.9, 27] the nuclei become stable when they reach a particular size (the critical nucleus). Above this size the nucleus is stable and continues to grow. However, in this case the size of the critical nucleus is of the order of 100 atoms or more. An alternative theory of nucleation has been proposed for critical nuclei consisting of only a few atoms [29.102, 103]. In this model certain geometries have been postulated for the critical nucleus under different conditions of supersaturation and substrate temperature. The rate of nucleation is given by

$$I = Ra_0^2 N_0 \left(\frac{R}{\gamma N_0} \right)^{n^*} \times \exp \left(\frac{(n^* + 1)Q_{ad} - Q_D + E_{n^*}}{k_B T} \right). \quad (29.43)$$

In this expression the symbols have the following meanings: R is the impingement rate of atoms on the substrate, a_0 is a characteristic jump distance of the atoms across the substrate surface, N_0 is the concentration of adsorption sites, $\gamma = h/k_B T$ is the vibrational frequency, where T is the substrate temperature, and n^* represents the

number of atoms in the critical nucleus. The remaining terms in the numerator of the exponential represent energies: Q_{ad} is the binding energy of a single atom to the substrate surface, Q_D is the activation energy for surface diffusion of adsorbed atoms and E_{n^*} is the dissociation energy of a critical nucleus containing n^* atoms. By substituting $n^* = 1, 2, 3$, etc., into (29.43), expressions can be obtained for critical nuclei of size 1, 2 or 3 atoms. These correspond to stable nuclei of size $(n^* + 1) = 2, 3$ or 4. By simultaneously solving pairs of these equations, it is possible to calculate the temperature at which the size of the critical nucleus changes from one value of n^* to another. For example, the temperature at which the critical nucleus size changes from one to two atoms, T_{12} , is given by

$$T_{12} = -\frac{(Q_{ad} + E_2)}{k \ln(R/\gamma N_0)}, \quad (29.44)$$

where E_2 is the dissociation energy of a two-atom nucleus into single atoms. The model has been tested, notably in the case of a face-centred cubic metal (silver) deposited on a high-purity (100) NaCl substrate [29.103]. At low substrate temperatures and high supersaturations, it was found that the nucleation rate I is proportional to R^2 , which from (29.43) implies that $n^* = 1$ (the critical nucleus is just one atom). At higher temperatures, critical nuclei were found to have $n^* = 2$ or 3, with stable nuclei adopting the (111) or (100) orientations. Further discussion is given in several later reviews [29.1, 10, 16, 19]. More-sophisticated models of

nucleation are based on the kinetic behaviour of nuclei, and involve a set of coupled differential equations, relating growth and dissociation of nuclei. More details are given in the literature [29.105, 106].

Another example of where the occurrence of epitaxy is defined by well-defined temperature ranges, is given in the work of Kalinkin et al. [29.107] Muravjeva et al. [29.104]. This important work has also been discussed in other reviews [29.108, 109]. These workers made extensive measurements on the evaporation of both cadmium and zinc chalcogenides (sulphides, selenides and tellurides), which were evaporated from Knudsen cells at a pressure of about 10^{-2} – 10^{-3} Pa. The substrates used were mica in all cases. It was found that for all three cadmium compounds, and all three zinc compounds, the conditions for epitaxial growth were dependent on both the temperature of the substrate (the epitaxial temperature T_{Ep}) and the evaporator temperature T_{ev} . The latter influenced the evaporation rate N_e (29.5) and also the supersaturation ratio of the growing films. For each of the six compounds, different combinations of T_{Ep} and T_{ev} were investigated. We discuss in detail only the cadmium compounds, although the behaviour of the zinc compounds is entirely similar, as can be seen from Fig. 29.11 [29.104]. Below about $T_{Ep} \approx 310$ – 320 °C, the evaporator temperature for epitaxy T_{ev} increased approximately linearly with T_{Ep} according to $T_{ev} = A_1 + T_{Ep}$, where A_1 is a constant. Above $T_{Ep} \approx 310$ – 320 °C, T_{ev} decreased linearly with T_{Ep} according to $T_{ev} = A_2 - 2T_{Ep}$, where A_2

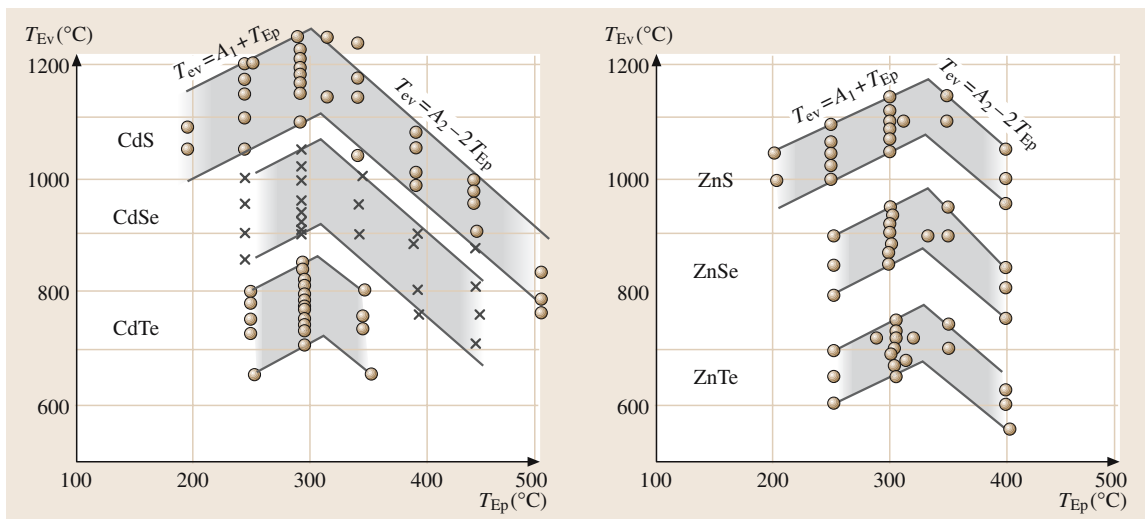


Fig. 29.11 Relationship between the evaporation temperature T_{ev} and the epitaxial temperature T_{Ep} in the growth of cadmium and zinc chalcogenide films on mica. After [29.104] with permission from Elsevier Science

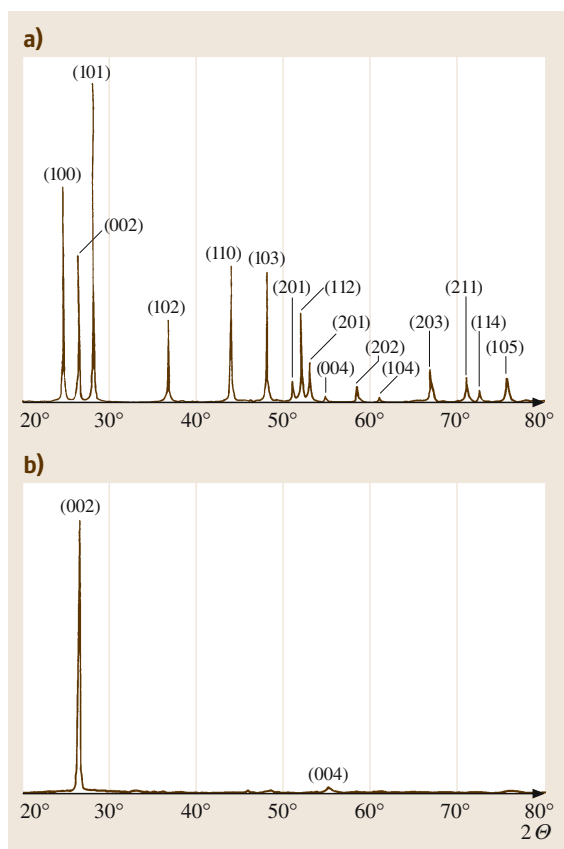


Fig. 29.12a,b X-ray diffraction traces obtained from (a) CdS powder, and (b) a CdS thin film. After [29.110] with permission from Wiley-VCH

is a second constant. These relationships between the epitaxial temperature and the evaporator temperature are illustrated in Fig. 29.11 [29.104]. For the cadmium chalcogenides, values of the range of the constants A_1 and A_2 for which epitaxial behaviour was observed are $A_1 = 800\text{--}950^\circ\text{C}$ and $A_2 = 1750\text{--}1900^\circ\text{C}$ for CdS, $A_1 = 600\text{--}750^\circ\text{C}$ and $A_2 = 1550\text{--}1700^\circ\text{C}$ for CdSe, and $A_1 = 400\text{--}550^\circ\text{C}$ and $A_2 = 1350\text{--}1500^\circ\text{C}$ for CdTe [29.104]. It is evident from these values that for each substance the evaporator temperature T_{ev} lies within a range of about 150°C for values of T_{Ep} at which epitaxy is possible. In addition, for a given substrate temperature the evaporator temperature required for epitaxial CdS is highest, becoming progressively lower for CdSe and for CdTe. For example for $T_{\text{Ep}} = 320^\circ\text{C}$, the evaporator temperatures for epitaxy are $1120\text{--}1270^\circ\text{C}$ for CdS, $920\text{--}1070^\circ\text{C}$ for CdSe and $720\text{--}870^\circ\text{C}$ for CdTe. The same sequence is also followed for the zinc

chalcogenides. At combinations of temperatures outside the limits given by the values of A_1 and A_2 , the films deposited were not epitaxial, being polycrystalline and only partially ordered.

We now consider some examples of polycrystalline thin films, and the factors which account for their various crystallographic characteristics. X-ray diffractometry is particularly useful in determining the basic crystal structure adopted by a particular film, particularly when two or more phases are possible. The cadmium chalcogenides have, on average, four valence electrons per atom, which leads to tetrahedral bonding. However, there are two different ways in which tetrahedra formed from each of the constituent atoms can penetrate into each other. This leads to two allowed crystal structures: the hexagonal wurtzite structure and the cubic zinc-blende structure. The planar spacings for these are described by (29.41) and (29.39), respectively. In the evaporant material, which is usually composed of a polycrystalline powder, a large number of diffraction peaks normally occur, which can be indexed using the appropriate planar spacing expression. Figure 29.12a [29.110] shows an example of a diffraction trace obtained from CdS evaporant powder. The diffraction planes are indexed in terms of the hexagonal structure, and rule out the possibility of the material being of the cubic form. More interestingly, Fig. 29.12b shows a trace obtained from a thin film prepared from this material. Although the two peaks shown correspond to those observed for the evaporant powder, it is clear that reflections from most of the crystal planes are missing. If the intensities of the peaks are different from those for a random powder (or more specifically from those tabulated in the *International Tables of Crystallography*), the crystallites are said to be preferentially oriented. In this case it can be seen that virtually all the crystallites must be aligned in the same direction, with the c -axis directed perpendicular to the substrate plane, along a preferential orientation in the $[002]$ direction.

Apart from the simple determination of film crystal structures and preferential orientations, it is also possible to monitor changes from one crystal structure to another. For example, the phthalocyanines are organic crystals, which are known to possess several basic crystal structures, of which the α -phase and the β -phase are the most common. There is some controversy over whether the α -phase is orthorhombic or monoclinic, and it was originally identified as tetragonal [29.111]. The β -phase is monoclinic. However, X-ray diffraction has identified most films prepared at lower substrate temperatures as being of the α -structure, while those deposited

at higher temperatures are of the β -structure. Both copper phthalocyanine (CuPc) and cobalt phthalocyanine (CoPc) films deposited on to substrates held at room temperature were of the α -structure, but on annealing at 240 °C for an extended period, the CuPc films underwent a phase transition to the β -structure [29.112]; the CoPc films also showed a transition to the β -phase on annealing at temperatures up to 325 °C [29.113]. The α -form films of both materials were preferentially aligned in the [001] direction assuming a tetragonal structure, or the [002] direction assuming an orthorhombic or monoclinic structure. The annealed β -form films were also preferentially oriented, in the [20 $\bar{1}$] direction for CuPc and the [001] direction for CoPc. In the latter case it was possible to monitor the dominant peak over the temperature range 200–325 °C where the phase change occurred. d_{hkl} increased slightly with annealing temperature up to 200 °C, then decreased significantly at 250–300 °C. The value of 2θ decreased slightly with T_{ann} up to 200 °C, again followed by a significant increase at approximately 250–300 °C. For T_{ann} up to 250 °C d_{hkl} lay in the range 1.271–1.277 nm and 2θ in the range 6.92–6.95°, which is close to those expected for α form CoPc; at higher annealing temperatures d_{hkl} was in the range 1.241–1.246 nm and 2θ in the range 7.09–7.12°, which are in better agreement with those expected for β form CoPc.

X-ray diffraction also allows an estimate of the mean grain size L in polycrystalline films to be obtained, by detailed observation of the peak profiles. The well-known Scherrer equation [29.114]

$$L = \frac{K\lambda}{\eta \cos \theta} \quad (29.45)$$

allows an estimate of L to be made from simple details of the peak profiles. In (29.45), λ represents the X-ray wavelength, η the width of a strong peak in radians at half-maximum intensity, and θ is the Bragg angle, where K is a constant which is approximately unity. An example of the use of this method is in the determination of L from the first Bragg peak of α -form CoPc films before annealing [29.113]. The films were subsequently heated to a higher temperature T_{ann} in oxygen-free nitrogen for two hours and the value of L again measured. Measurements were repeated at even higher temperatures until the dependence of L on T_{ann} over the entire temperature range of the α -to- β phase change was determined. These results are illustrated in Fig. 29.13 [29.113]. It is clear that at temperatures up to 200 °C there was an increase in the mean microcrystallite size from about 30 nm to over 40 nm, although there was no indication of a phase

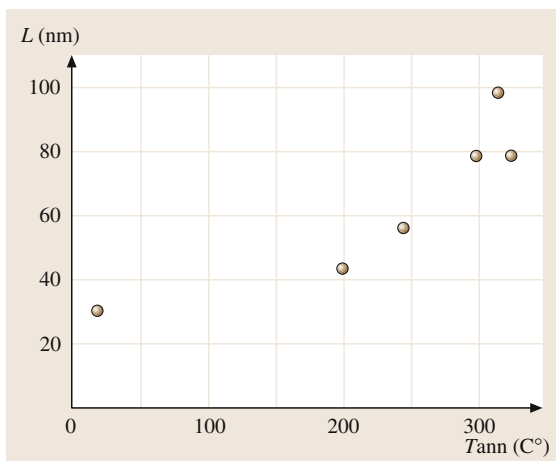


Fig. 29.13 Mean crystallite grain size L as a function of maximum cumulative annealing temperature T_{ann} for a cobalt phthalocyanine film of thickness 0.1 μm . After [29.113] with permission from Wiley-VCH

change. Ashida et al. [29.115] ascribed this effect in CuPc to preliminary crystallite growth at the early stages of the phase transformation process. Above annealing temperatures of about 200 °C there was a rapid increase in crystallite size to about 100 nm at 315 °C; similar behaviour has also been observed in which α -form CuPc films were heated to 300 °C when the diffraction patterns became sharper and the crystallites larger [29.115].

Some general conclusions concerning the dependence of grain size in vapour-deposited polycrystalline films on various film attributes have been summarised by Chopra [29.18]. The major feature determining mean grain size is the surface mobility of adsorbed atoms and clusters during the deposition process. A high surface mobility will allow adsorbed atoms to move large distances before being incorporated into growing nuclei, resulting in a relatively low concentration of nuclei and thus larger crystallites. The substrate temperature is the most obvious factor contributing to high mobility, and therefore grain size tends to increase with increasing substrate temperature. We have already seen that epitaxy requires a minimum substrate temperature; this too is a consequence of the surface mobility requirement. The total film thickness obviously sets an upper limit on the crystallite size, and therefore the grain size increases in general with film thickness. Subsequent annealing at temperatures greater than the substrate temperature during deposition can also cause an increase in the mean crystallite size as was shown in Fig. 29.13, as surface atoms augment the crystallites and grains coalesce.

The dependence of grain size on deposition rate is not consistent between different materials and substrates. Frequently there is an increase with increasing deposition rate [29.18], although a decrease has been observed in some cases, such as in CdTe films where the establishment of a higher concentration of adsorbed nuclei led to the growth of finer grains [29.38].

The interplay between the various deposition parameters can have a profound effect on the structure of the films. This is illustrated particularly well in Fig. 29.14 [29.116], for cadmium arsenide (Cd_3As_2) films evaporated onto mica substrates. Both the substrate temperature T_s and the growth rate v were varied, with the latter covering a wide range. Diffraction patterns indicated that the crystalline films were of the lower temperature tetragonal α -phase of Cd_3As_2 ; the structure of this phase is documented in the literature [29.117] and the planar spacings are given by (29.40). Amorphous films recrystallised following interaction with the electron beam in the electron microscope and also reverted to a tetragonal structure. The figure shows that, over the substrate temperature and deposition rate ranges investigated, the films could have an amorphous, polycrystalline or well-oriented crystal structure. Below approximately 125–145 °C the films were amorphous, with polycrystalline films resulting when this temperature was exceeded. The amorphous-to-crystalline transition was only weakly dependent on the deposition

rate. However, the epitaxial temperature, which delineates polycrystalline and epitaxial films, increased with increasing deposition rate. For deposition rates less than about 4 nm/s (40 Å/s) there was a direct transition from the amorphous to the well-oriented crystalline structure, whereas at higher rates there was a transitional polycrystalline region. The preferred orientation of the films was [001] with the c -axis of the tetragonal structure perpendicular to the plane of the substrate. Also shown on the figure are examples of X-ray diffraction patterns for the crystalline, polycrystalline and amorphous regions. These are characterised, respectively, by discrete reflection points, *spotty* rings and diffuse rings.

Finally, a brief discussion on amorphous films is appropriate. These are films which show no evidence of a regularly repeated crystalline structure, i.e. they do not exhibit long-range order. We have already seen that in Cd_3As_2 films (Fig. 29.14), a relatively low substrate temperature during deposition leads to the growth of amorphous films. This is a general rule, applicable to most materials, and is basically the opposite to the condition for epitaxy – the deposition of amorphous films is encouraged if the surface mobility of atoms on the substrate is low. This causes the adsorbed atoms to be incorporated into the films almost at the point of impact, rather than after diffusion across the substrate surface and joining a growing nucleus. High deposition rates also predispose atoms to remain at their impact position

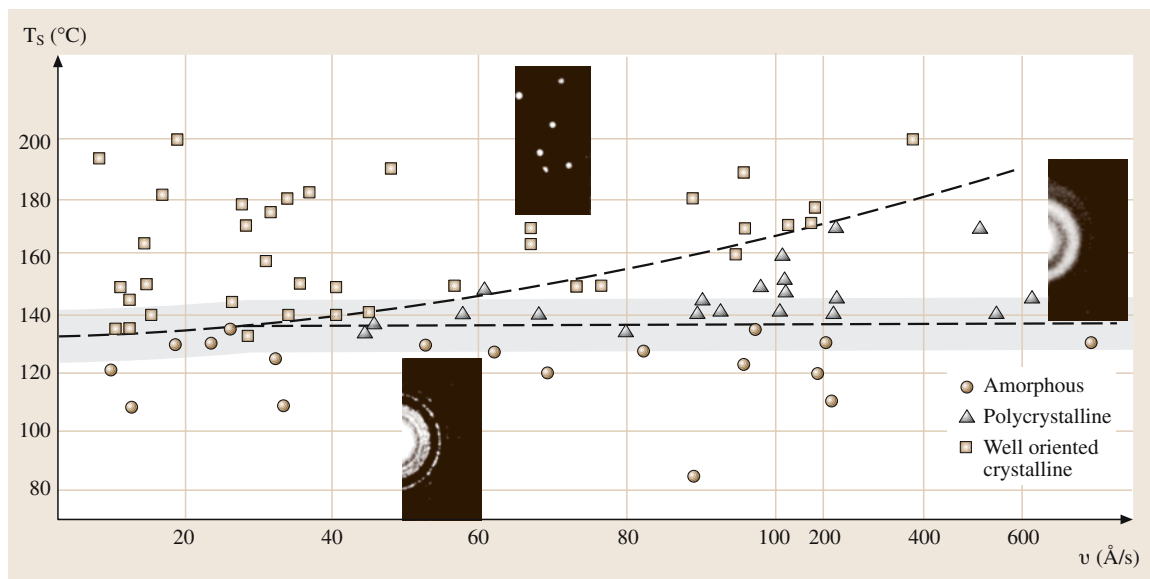


Fig. 29.14 Effects of film growth rate v and substrate temperature t_s on the structure of Cd_3As_2 films vacuum-evaporated onto mica substrates. After [29.116] with permission from Elsevier Science

within the growing film – they have only a very limited time to diffuse before they are locked into the growing film structure. These considerations are pertinent to the deposition of amorphous metallic films, which require substrate temperatures to be lowered to liquid-nitrogen temperature or below, and also to elemental chalcogenide and semiconductor films (Ge and Si) and carbon, all of which have a sufficiently low surface mobility at room temperature. A table giving examples of amorphous thin film materials, their preparation methods, and the transition from the amorphous to a crystalline phase is given by *Maissel and Francombe* [29.16]. The amorphous state is metastable, requiring some thermal energy to transfer it to a crystalline state.

Another method of producing amorphous films is the incorporation of residual gases during deposition. A relatively high pressure of oxygen ($\approx 10^{-3}$ Pa) can cause oxidation of the evaporant. Large-scale crystallites of metals are then unable to form. The admission of nitrogen during sputtering of refractory metals is also sufficient to result in the deposition of an amorphous phase [29.1].

Since the structure of amorphous materials do not possess any long-range order, it is necessary to describe the short-range order using some function which can be determined directly from diffraction experiments. The radial distribution function $J(r)$ is defined as the number of atoms lying at a distance between r and $r + dr$ from the centre of an arbitrary atom at the origin [29.119]. It is given by

$$J(r) = 4\pi r^2 \rho(r). \quad (29.46)$$

Here, $\rho(r)$ is an atomic pair correlation function, which is equal to zero at values of $r < r_1$, where r_1 is the average nearest neighbour interatomic separation, $\rho(r)$ is equal to the average value of density ρ^0 at large values of r where the material becomes homogeneous. Between these values $\rho(r)$ is oscillatory, with the peaks in the distribution corresponding to the average interatomic separations [29.119]. $J(r)$ also shows oscillations superimposed on the average density parabola $4\pi r^2 \rho^0$, and the area under a particular peak gives the co-ordination number for the given shell of atoms. The coordination number represents the number of nearest neighbours in the structure. An example of a radial distribution function, derived from X-ray data, for thin films of the compound $\text{Ge}_x\text{Te}_{1-x}$ where $x = 0.11$, is shown in Fig. 29.15 [29.118]. For materials having covalent bonds the coordination number is four, whereas for amorphous metals, which generally crystallise in close-packed structures, the coordination number is higher.

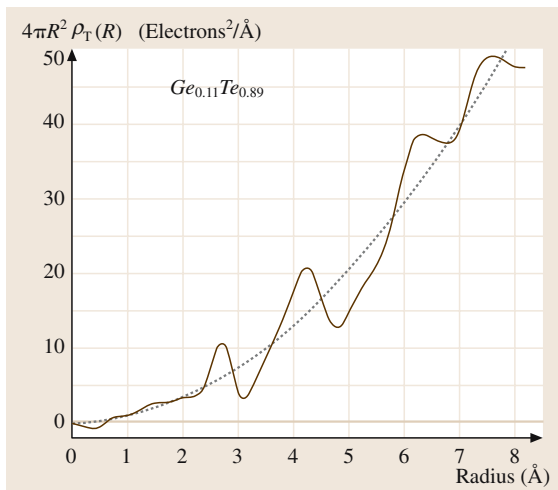


Fig. 29.15 Calculated radial distribution function for a sample of composition $\text{Ge}_{0.11}\text{Te}_{0.89}$. The dashed curve represents the contribution of the average electron density. After [29.118] with permission from Elsevier Science

For covalently bonded materials, the atoms tend to arrange themselves in tetrahedra, which in the amorphous state interconnect at their apexes. However, the apexes themselves interconnect within the material in a random manner, whereas the tetrahedra dimensions are effectively determined by the size of the atoms. Thus there is no long-range order because of the random interconnections, but the short-range order survives, and is similar to that for the corresponding crystalline material. *Bosnell* [29.120] has described the important amorphous semiconducting thin films, and classified them into three major categories: covalent noncrystalline solids, semiconducting oxide glasses and dielectric films. The importance of these materials, in particular their electrical properties, cannot be overestimated. The standard text on this subject [29.121] gives copious information on their electrical properties, as well as their structures and radial distribution functions. Local atomic order in amorphous films is discussed in the literature by *Dove* [29.122], with the primary emphasis being on data derived from electron diffraction. This work also gives the theory behind the radial distribution function, and provides a wide review of radial distribution functions measured in many systems performed prior to the date of the review.

From the foregoing discussion and examples, it should be clear that the crystallographic structure of thin films, whether epitaxial, polycrystalline or amorphous, depends heavily on the deposition conditions. In turn

the film structure strongly influences the mechanical, thermal, optical and electronic properties.

29.2.3 Morphology

Perhaps the most striking morphological characteristic observed in thin films grown by the physical vapour deposition techniques is that of the columnar grain structure. Electron microscopy frequently reveals that high-density regions tend to grow in parallel columns. They tend to develop when the surface mobility of adsorbed atoms is relatively low, so that the area of the base of the columns is small. Further growth of the grains is away from the substrate. The direction of column growth is not necessarily perpendicular to the substrate surface, although it frequently is for depositing material arriving at normal incidence. For oblique incidence it has been observed that the columns are oriented away from the substrate normal and towards the vapour source, although not necessarily pointing directly towards it. If α is the angle between the substrate normal and the vapour stream, and β is the angle between the substrate normal and the direction of the columns, it has been shown experimentally that β is usually less than α . The tangent rule [29.124] relates these two angles and is given by

$$\tan \alpha = 2 \tan \beta \quad (29.47)$$

and was found to be followed for the case of directly evaporated aluminium. In experiments on CdS films, Hussain [29.125] deposited films of thickness up to 100 μm with angle of incidence α varying between 0 and 60°. Up to a certain film thickness the c -axis of the grains was perpendicular to the direction of vapour incidence, but above this thickness the c -axis started to shift towards the substrate normal, eventually becoming aligned with it. Etching of a thin film deposited at oblique incidence confirmed that the top layers of the film had their c -axis normal to the substrate, whereas in the bottom layers it was parallel to the direction of vapour incidence. In other work on this material [29.126] similar thickness-dependent orientation effects have been reported. In this case, films of thickness a few tens of nanometres had randomly orientated crystallites, while those of thickness of about 20 μm were aligned with the c -axis normal to the substrate.

Evaporated CdS films also show an interesting morphology on a somewhat larger scale, which is illustrated in Fig. 29.16 [29.123]. This illustrates features which are evident in many different types of thin films. The top surface of as-deposited films of thickness 25–30 μm were observed by scanning electron microscopy to consist of

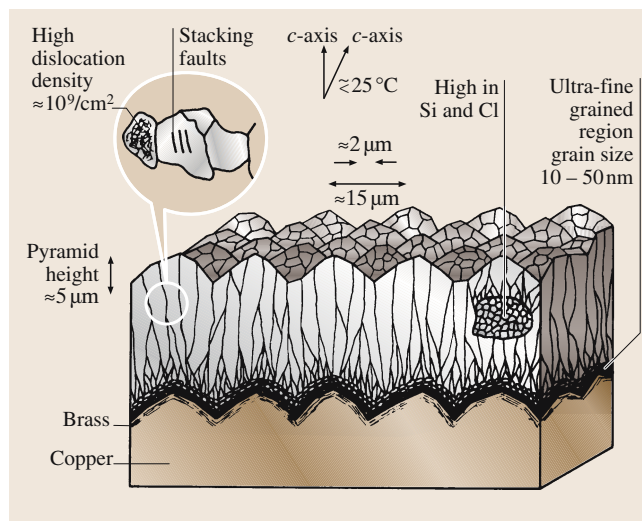


Fig. 29.16 Schematic representation of evaporated CdS film microstructure inferred from scanning electron microscopy studies. After [29.123] with permission from Elsevier Science

partially faceted hills with a height of 5 μm and a width of about 15 μm at the base. These workers also determined the grain size and crystal defect structure using transmission electron microscopy. The grain size was about 2 μm at the top and at the centre of the films, but was only of the order of 10–50 nm at the back surface adjacent to the substrate. Typical dislocation densities were of the order of 10^{13} m^{-2} (10^9 cm^{-2}). Ultrafine-grain regions of 2–15 μm in diameter were observed. These were thought to be spits of material rapidly ejected from the source. Similar features have also been observed in this laboratory [29.110]. The partially faceted protrusions were reported, and cracks in the films were also present, as a result of differential thermal contraction between the CdS films and the glass substrates during cooling. Larger regions of CdS were seen, which were considered to have been ejected from the evaporation boat during deposition. This phenomenon has been referred to as *spattering* or *splattering* by Stanley [29.127], and its elimination is an important goal for deposition processes operating at higher temperatures.

In early work on CdTe films [29.128] photovoltages were measured when there was an angle between the vapour stream and the normal to the substrate. This led to the suggestion that in this case the film growth was directionally anisotropic where the nucleating material depositing on a heated substrate has sufficient surface mobility to form islands rather than a uniform thin film. Additional material arrives at an angle, shadowing the

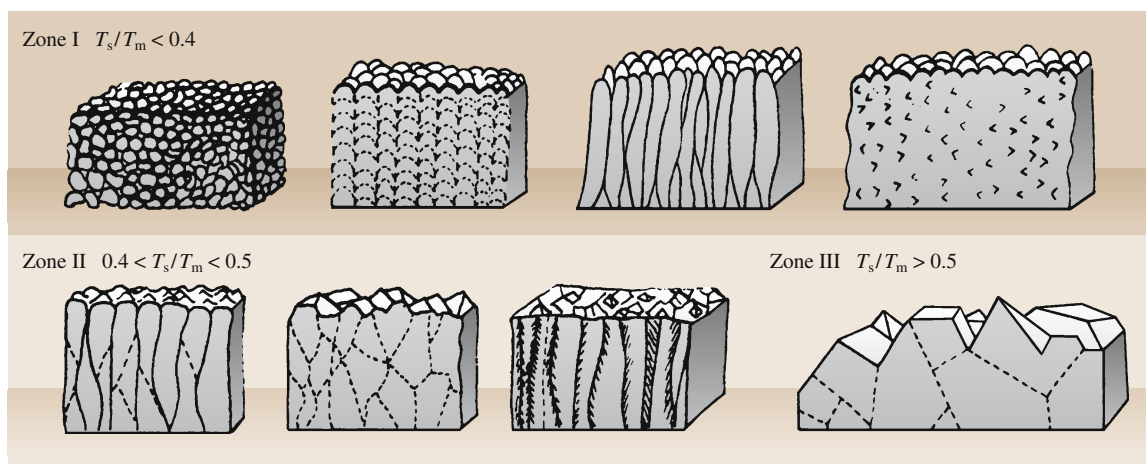


Fig. 29.17 Schematic three-zone model for the temperature dependence of the growth morphology in evaporated Cd_3As_2 films. T_s/T_m represents the ratio of the substrate temperature to the melting point of Cd_3As_2 . After [29.131] with permission from Elsevier Science

far side of the island from the deposition source, and the shadowed side grows relatively slowly and is well ordered, while the exposed face grows towards the vapour source. When islands begin to touch each other there is expected to be a mismatch, leading to a disordered state in the region between the crystallites. Each interface will therefore have an ordered and a disordered side and at the interface between grains a photovoltage develops. Photomicrographs of epitaxial CdTe layers grown by the close-space technique [29.129] revealed smooth epitaxial layers when the deposition was on to (100), (110) and (111) Cd oriented substrates, but large hillocks grew on (111) Te substrate faces. When grown on mica substrates [29.130] it was observed that the morphology was significantly affected by the composition of the incident vapour phase. This was characterised by the ratios γ_{Cd} and γ_{Te} , which depended on the ratio of the actual concentration to the stoichiometric concentration of the Cd atoms or the Te_2 molecules in the gas phase. When $\gamma_{\text{Cd}} = 53$ when CdTe and Cd were co-evaporated the films were of high quality consisting of large crystallites with degenerated (111) faces. When CdTe and Te were co-evaporated and $\gamma_{\text{Te}} = 12$ formations resembling dendrites were observed.

CdSe film morphologies have been observed by several workers. Tanaka [29.132] prepared films using Ar and H_2 sputtering gases; some films were subsequently annealed in Ar . Microscopic examination showed that sputtering in H_2 gas led to uniform grain growth of the films over the whole substrate surface, while films sputtered and annealed using Ar gave rise to geometric

pillars and a textured structure. An investigation of the optimum conditions for obtaining CdSe epitaxial layers on (0001) oriented sapphire substrates was performed by Ratcheva-Stambolieva et al. [29.133]. For substrate temperatures up to 460°C the layers were polycrystalline, from 460°C to 530°C a small degree of preferred orientation was observed, with a more significant textured structure for temperatures of 530 – 570°C . At higher temperatures of 570 – 620°C a monocrystalline structure was obtained, varying in its degree of perfection depending on the actual temperature and the growth rate. Hexagonal pyramids with peaks, regular or irregular facets and flat tops were also identified. RF-sputtered CdSe films [29.134] were porous when deposited at a substrate temperature of 550°C onto vitreous silica. When deposited onto substrates of single-crystal sapphire the surface appeared shiny, and showed features of hexagonal flat tops and pyramids with edge lengths in the range 1 – $5\text{ }\mu\text{m}$.

We finally look at the morphological characteristics of another cadmium compound, Cd_3As_2 . Evaporated films of this material have been investigated by Jurusik and Żdanowicz [29.131] using electron microscopy. It was found that the microstructure of the films was strongly dependent on the substrate temperature T_s . This type of behaviour is frequently addressed in terms of a zone model, and examples of some of these for other materials are discussed by Ohring [29.19]. In Cd_3As_2 the microstructures were interpreted in terms of a simple three-zone model in which the structure was correlated with the ratio T_s/T_m , where $T_m = 994\text{ K}$

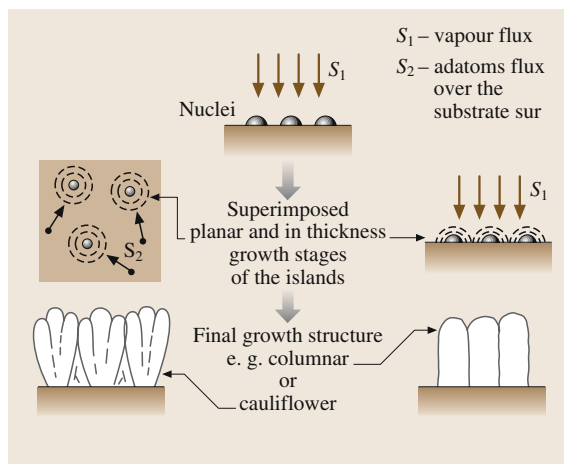


Fig. 29.18 Post-nucleation growth in evaporated Cd_3As_2 films considered as a consequence of two superimposed (planar and thickness) growth stages. After [29.135] with permission from Elsevier Science

is the melting point of Cd_3As_2 [29.131]. They found that for $T_s/T_m \leq 0.4$ the films were amorphous, consisting of spherical or elongated fibrous-like clusters, or were uniform with a large number of pores or cavities. For $0.4 < T_s/T_m < 0.5$ columnar growth was identified, while for $T_s/T_m \geq 0.5$ films with large crystallites and clear grain boundaries were observed. These structures are illustrated in Fig. 29.17 [29.131]. More recently a very detailed study was performed on amorphous Cd_3As_2 films evaporated onto substrates maintained at 293 K [29.135, 136]. It was found that the size of post-nucleation islands was a function of the substrate temperature, type of substrate and the deposition rate. Furthermore the shape and type of growth of the islands varied with these parameters. At a low deposition rate (1 nm/s) the initial nuclei grew until a certain size was reached, and new nuclei started to grow around the original islands. For a moderate deposition rate (3.5 nm/s) the initial nuclei grew into islands which in turn increased in size until the islands had only narrow interfaces between them. Some islands formed polygons. Films deposited at a high rate (50 nm/s) consisted of small islands connected by relatively large low-density areas. Films deposited at low rates were of approximately stoichiometric proportions (i. e. 40 at. % As) while those deposited at higher rates had excess arsenic (typically 51.8 at. % As). It was suggested that the growth process could be considered as a superposition of two growth stages, as shown in Fig. 29.18 [29.135]. The vapour flux onto the substrate is denoted by S_1 and the

adatom flux over the substrate surface by S_2 . In the first stage of growth the initial nuclei enlarge owing to the adatom flux S_2 . The second stage of growth incorporates atoms immediately from the incident flux S_1 . This is usually very slow, but increases as the island surface area increases. At this stage branched complex islands or only compact islands grow as the result of a shadowing mechanism. When the islands touch each other the molecular flux S_2 over the substrate surface is eliminated, and the films grow in thickness as the result of the incident flux S_1 and the structure initiated in the earlier stage (i. e. large branched or smaller compact islands). The final growth structure gives rise to cauliflower-like or columnar structures. In later work [29.136] Cd_3As_2 films were evaporated from bulk crystallites of Cd_3As_2 or CdAs_2 . For films grown from the Cd_3As_2 source a pronounced columnar structure was observed, whereas those evaporated from the CdAs_2 source grew uniformly, and the columnar structure was absent. In the latter case the initial island structure was not observed in the early stages of film growth. It was concluded that prerequisites for columnar growth were a limited surface mobility and an oblique component in the deposited flux, particularly where the films are deposited normally to the substrate surface.

In this section some of the morphological characteristics of some cadmium compound films prepared by PVD methods have been described. The emphasis in this discussion of the films was not primarily because of their usefulness as semiconductors in various applications, but because they exhibit many features typical of a wide range of other film materials. Surface mobility of atomic species on the substrate during deposition is clearly of great importance, as this tends to control the grain size at the substrate interface, while the angle of vapour incidence relative to the substrate normal is an important factor in determining the major direction of crystallite growth and whether a columnar structure develops. Shadowing of the vapour stream by islands during film growth can lead to differential growth rates at different points on the film surface. Bulk diffusion and desorption also contribute to the film morphology variations [29.19]. The morphological characteristics of many materials may be explained using zone models in which the substrate temperature and other variables, such as sputtering gas pressure, conspire to determine the overall morphology via the mechanisms detailed above. The morphology of thin films is particularly rich, and an important factor in influencing film properties.

29.3 Properties

As mentioned in the introduction, the properties covered in this section are restricted to the optical and electrical properties. Moreover, only the very fundamental properties, which are a direct consequence of the thin-film nature of the material, are covered. Thus the more advanced features, such as those exhibited by specific classes of materials, individual elements or compounds, and electronic, optical and optoelectronic devices are not normally discussed, although many of these are considered elsewhere in this Handbook. The basic thin-film properties, which have been exploited in many of these developments are stressed. Examples include the use of thin films in layers or stacks in anti-reflection coatings and mirrors [29.30], the contribution of grain boundaries in conducting thin films to the resistivity of tracks and connections in integrated circuits, the importance of high-field conduction in some electronic devices [29.17], and the possible influence of high-field conduction in nanostructures [29.11].

29.3.1 Optical Properties

This section comprises an introduction to the main optical properties of thin films, starting with a discussion of the properties of nonabsorbing single films at normal and non-normal incidence, and then considering modifications due to absorption. Finally a brief discussion of the use of multilayer film stacks is given. Thin-film optics is a particularly wide field, and readers interested in further details and a more in-depth analysis are referred to one of the standard texts [29.30, 137, 138], or indeed to most general textbooks on optics, nearly all of which cover thin films in some detail. As mentioned earlier, optical effects in thin films were first recorded in the 17th century, which predates by two centuries the emergence of Maxwell's equations, which encapsulate the behaviour of all electromagnetic radiation, including visible light. The boundary conditions at the interface between two different optical media are derivable directly from Maxwell's equations, and allow us to calculate the reflection and transmission coefficients for the amplitude of the electromagnetic wave at the interface. These are termed the Fresnel coefficients.

Figure 29.19a shows the simplest situation, where a beam of light in a medium having refractive index n_0 is incident at an angle φ_0 to the normal on a film of refractive index n_1 . The following expressions for the Fresnel coefficients apply to all individual interfaces in a structure, and for all values of the refractive index. At

this stage in the discussion we assume that the materials are isotropic and nonabsorbing. At the interface the electromagnetic radiation is refracted into the medium of index n_1 at an angle φ_1 , with a reflected component directed back into the incident medium at an angle φ_0 . Conventionally, the coefficients are quoted for components of the light beam resolved parallel (denoted by

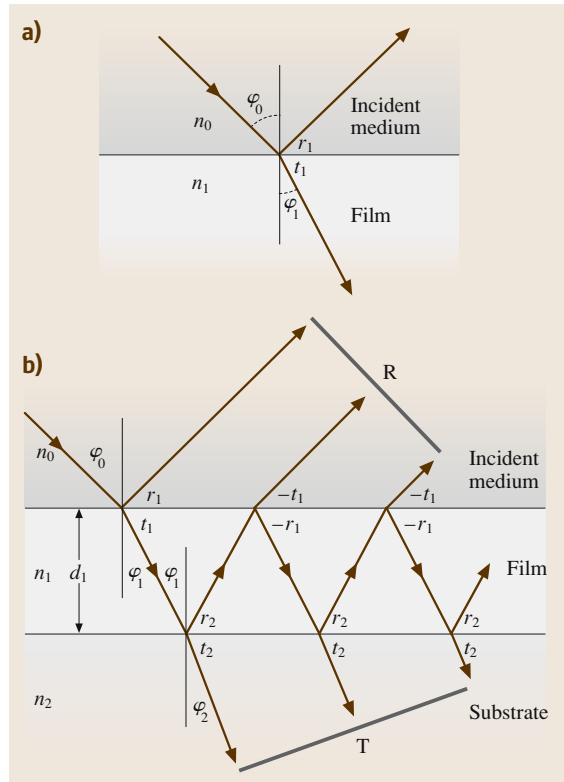


Fig. 29.19 (a) Reflection and transmission at a single interface between two media with refractive indices n_0 and n_1 . r_1 and t_1 represent the amplitude reflection and transmission coefficients at the interface (the Fresnel coefficients). (b) Total reflection and transmission as the result of multiple reflection and transmission through a nonabsorbing thin film of thickness d_1 and refractive index n_1 deposited on a substrate of refractive index n_2 . The Fresnel coefficients are indicated at each transit of a boundary, and change sign if the direction of incidence is reversed. R and T represent the total amplitude of reflectance and transmittance, respectively, each obtained by summing the amplitudes of an infinite number of components, although only three are shown

p) and perpendicular (denoted by s) to the plane of incidence. Therefore there are two reflection coefficients r_{1p} and r_{1s} , and two transmission coefficients, t_{1p} and t_{1s} . The 1 in the subscripts in this notation refers to the first interface, i. e. between the materials with refractive indices n_0 and n_1 in this case. More complex systems will have more than one interface, where 2, 3, etc., signify the second or third interfaces. The Fresnel coefficients for reflection are [29.1, 15]

$$\begin{aligned} r_{1p} &= \frac{n_1 \cos \varphi_0 - n_0 \cos \varphi_1}{n_1 \cos \varphi_0 + n_0 \cos \varphi_1}, \\ r_{1s} &= \frac{n_1 \cos \varphi_1 - n_0 \cos \varphi_0}{n_1 \cos \varphi_1 + n_0 \cos \varphi_0}, \end{aligned} \quad (29.48)$$

and those for transmission are

$$\begin{aligned} t_{1p} &= \frac{2n_0 \cos \varphi_0}{n_1 \cos \varphi_0 + n_0 \cos \varphi_1}, \\ t_{1s} &= \frac{2n_0 \cos \varphi_0}{n_1 \cos \varphi_1 + n_0 \cos \varphi_0}. \end{aligned} \quad (29.49)$$

Clearly these coefficients reduce to

$$r = \frac{n_1 - n_0}{n_1 + n_0}, \quad t = \frac{2n_0}{n_1 + n_0} \quad (29.50)$$

for the case of normal incidence where all the cosine terms become unity. There is no difference in this case between the coefficients for the p and s components. An expression for the energy reflectance at the interface R can be obtained by evaluating rr^* , where r^* is the complex conjugate of r . For nonabsorbing films r is real and this reduces to r^2 , and gives for normal incidence

$$R = \left(\frac{n_1 - n_0}{n_1 + n_0} \right)^2. \quad (29.51)$$

A similar expression for the energy transmittance T at the boundary can also be derived from t , but in this case the incident and transmitted beams are in different media, and the expression is not simply r^2 , but includes the ratio n_1/n_0 . Full details are given elsewhere [29.18, 138].

We consider now the slightly more complex situation of a thin film of thickness d_1 and refractive index n_1 deposited on a substrate of refractive index n_2 . In this case the value of n_0 for the incident medium will usually be the free-space value of unity. There will be multiple reflections as shown in Fig. 29.19b. It is clear that the total amplitude of the reflected beam is given by summing the amplitudes of the individual reflections into the incident medium, while the total amplitude of the transmitted beam is obtained by summing the amplitudes of the transmitted components into the substrate.

In traversing the film of thickness d_1 , the radiation undergoes a change in phase, which is accounted for by a *phase thickness* δ_1 , given by

$$\delta_1 = \frac{2\pi}{\lambda} n_1 d_1 \cos \varphi_1, \quad (29.52)$$

where λ is the wavelength in vacuum. In each case the summations lead to infinite series. These involve $\exp(-i2\delta_1)$ for reflection, where the beam traverses the film an even number of times, and $\exp(-i\delta_1)$ for transmission, where the beam traverses an odd number of times. The results of these summations give the amplitude reflectance R and transmittance T , respectively, as

$$R = \frac{r_1 + r_2 \exp(-2i\delta_1)}{1 + r_1 r_2 \exp(-2i\delta_1)} \quad (29.53)$$

and

$$T = \frac{t_1 t_2 \exp(-i\delta_1)}{1 + r_1 r_2 \exp(-2i\delta_1)} \quad (29.54)$$

In this case r_1 and t_1 represent the Fresnel coefficients between n_0 and n_1 , and r_2 and t_2 the coefficients between the media n_1 and n_2 . Depending on whether we are considering parallel or perpendicular polarisation (p or s) the relevant Fresnel coefficients from (29.48) and (29.49) should be used in these expressions. Equation (29.53) and (29.54) refer to the ratios of the *amplitudes* of the reflected and transmitted beams to that of the incident beam. In terms of the beam *energies*, the reflectance R and the transmittance T may be defined as the ratios of the reflected and transmitted beam energies to that of the incident energy. The energies of the reflected and transmitted beams are given by $n_0 R R^*$ and $n_2 T T^*$ respectively, where R^* and T^* represent the complex conjugates of R and T . We then obtain

$$R = \frac{n_0}{n_0} R R^* = \frac{r_1^2 + r_2^2 + 2r_1 r_2 \cos 2\delta_1}{1 + r_1^2 r_2^2 + 2r_1 r_2 \cos 2\delta_1} \quad (29.55)$$

for the reflectance, and

$$T = \frac{n_2}{n_0} T T^* = \frac{n_2}{n_0} \frac{t_1^2 t_2^2}{(1 + 2r_1 r_2 \cos 2\delta_1 + r_1^2 r_2^2)} \quad (29.56)$$

for the transmittance. Although these latter two expressions are reasonably compact, when the four Fresnel coefficients are substituted they become somewhat unwieldy. However, for the important case of normal incidence, where (29.50) applies, and where (29.52) simplifies to $\delta_1 = (2\pi/\lambda)n_1 d_1$, they reduce to still fairly

long expressions, but involving only the three refractive indices and δ_1 , i. e.

$$R = \frac{(n_0^2 + n_1^2)(n_1^2 + n_2^2) - 4n_0n_1^2n_2 + (n_0^2 - n_1^2)(n_1^2 - n_2^2) \cos 2\delta_1}{((n_0^2 + n_1^2)(n_1^2 + n_2^2) + 4n_0n_1^2n_2 + (n_0^2 - n_1^2)(n_1^2 - n_2^2) \cos 2\delta_1)} \quad (29.57)$$

and

$$T = 8n_0n_1^2n_2 / [(n_0^2 + n_1^2)(n_1^2 + n_2^2) + 4n_0n_1^2n_2 + (n_0^2 - n_1^2)(n_1^2 - n_2^2) \cos 2\delta_1] \quad (29.58)$$

The variations of R and T with δ_1 are shown in Fig. 29.20 [29.139] for normal incidence, where the refractive index of the incident medium $n_0 = 1$ and that of the substrate $n_2 = 1.5$. These correspond to air and glass respectively. Similar plots of these variations as a function of optical thickness n_1d_1 expressed as multiples of the wavelength λ are also common in the literature [29.1, 16, 138], where an optical thickness of unity corresponds to $\delta_1 = 2\pi$. The figure shows curves for three values of the film refractive index of 2.00, 1.70 and 1.23. For the first two of these $n_0 < n_1 > n_2$ and maxima in reflectance occur at $\pi/2$ and $3\pi/2$ with minima at $0, \pi$ and 2π , while for the third $n_0 < n_1 < n_2$ and the positions of the maxima and minima are reversed. In general the maximum and minimum reflectance values R_{\max} and R_{\min} , depend on the relative values of n_0, n_1 and n_2 . For the film index greater or less than those at both its boundaries, i. e. $n_0 < n_1 > n_2$ or $n_0 > n_1 < n_2$, we obtain

$$R_{\max} = \left(\frac{n_1^2 - n_0n_2}{n_1^2 + n_0n_2} \right)^2, \quad R_{\min} = \left(\frac{n_2 - n_0}{n_2 + n_0} \right)^2 \quad (29.59)$$

with maxima occurring when δ_1 is an odd multiple of $\pi/2$ (optical thickness n_1d_1 an odd number of quarter-wavelengths) and minima when δ_1 is an even multiple of $\pi/2$ (n_1d_1 an even number of quarter wavelengths). For the film index intermediate between those at its boundaries, i. e. $n_0 < n_1 < n_2$ or $n_0 > n_1 > n_2$ these results are reversed, and we obtain

$$R_{\max} = \left(\frac{n_2 - n_0}{n_2 + n_0} \right)^2, \quad R_{\min} = \left(\frac{n_1^2 - n_0n_2}{n_1^2 + n_0n_2} \right)^2 \quad (29.60)$$

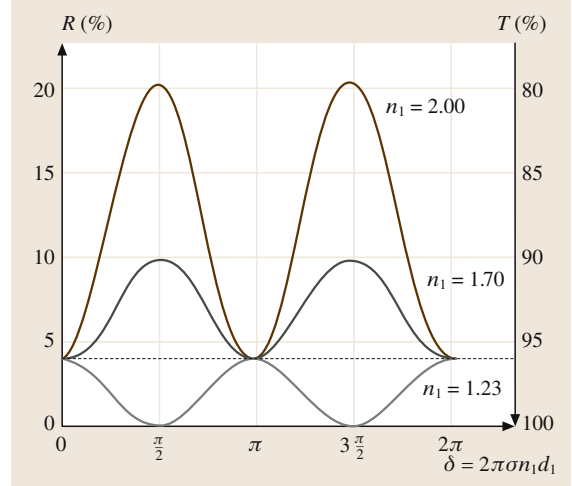


Fig. 29.20 Variation of the energy reflectance R and transmittance T with phase thickness δ_1 for a nonabsorbing film at normal incidence for various film indices n_1 , with $n_0 = 1$ and $n_2 = 1.5$. The wavenumber σ is the reciprocal of the wavelength in vacuum λ . After [29.139] with permission from Elsevier Science

with maxima occurring when δ_1 is an even multiple of $\pi/2$, and minima at odd multiples (or maxima for n_1d_1 an even number of quarter wavelengths and minima an odd number). Equation (29.60) predicts that zero reflectance occurs when $n_1 = \sqrt{n_0n_2}$, and for the refractive index values of Fig. 29.20 give $n_1 = 1.23$. Thus a simple *anti-reflection coating* can be made by using a film of this index with optical thickness $\lambda/4$. This is clearly evident from the figure at the points where δ_1 has values of $\pi/2$ and $3\pi/2$. Another point which emerges from the figure is that if the refractive index of the film is the same as that of the glass, the system behaves as if the glass is uncoated. This situation is shown by the dotted line in the figure. Here the reflectance is 4%, exactly as predicted by (29.51) for an uncoated substrate (i. e. $n_0 = 1$ and $n_1 = 1.5$).

If the electromagnetic light wave is attenuated by a film layer, that layer is said to be *absorbing*, in contrast to the *transparent* layers considered previously. Providing the incident light beam is incident *normally* on the surface of the absorbing medium, the theory described above may be applied, with the proviso that the refractive index n is replaced by a *complex refractive index* $\mathbf{n} = n - ik$. The quantity k is called the *extinction coefficient*, and characterises absorption of energy within the film. Substitution of the complex refractive index in the appropriate expressions for the Fresnel coefficient

of (29.50), and in the expression for the phase thickness of (29.52) at normal incidence allows the amplitude reflectance R to be obtained. The energy reflectance R can be obtained by substituting the complex expressions for r and δ directly into (29.53) and evaluating RR^* . For more than two or three layers the results become intensely unwieldy, and matrix methods as mentioned below are considerably more useful. A simple example of the use of the complex refractive index will suffice. If we take the case of an incident beam in a nonabsorbing medium such as air with refractive index n_0 , incident upon an absorbing medium with complex refractive index $n_1 = n_1 - ik_1$, the amplitude reflectance can be determined by replacing n_1 by $n_1 - ik_1$ in (29.50) and evaluating rr^* . In this case the energy reflectance at the boundary becomes

$$R = \frac{(n_0 - n_1)^2 + k_1^2}{(n_0 + n_1)^2 + k_1^2} \quad (29.61)$$

in contrast to the result of (29.51) for a nonabsorbing medium.

In general, in absorbing films, the planes of constant phase are not parallel to the planes of constant amplitude for the electromagnetic wave disturbances. Only for the special case of normal incidence is this true, and thus in absorbing films it is usual to consider normal incidence only. The general case of oblique incidence is discussed in the standard texts on the optics of thin films [29.137, 138]. The behaviour of the energy reflectance and transmittance depend on the optical constants n and k . We have seen that in transparent films, R and T oscillate with the film thickness due to interference effects. In absorbing films with small k values, the oscillatory behaviour remains, but the amplitudes of the maxima decrease with increasing films thickness. For higher absorption values, particularly in the case of metals, intensity maxima are not observed, and an exponential decrease in transmittance with film thickness is seen.

Of particular use in the case of thin film stacks containing several layers is the matrix formalism, which allows the reflection and transmission properties to be determined. A 2×2 matrix characterises each layer of the stack, and allows the electric and magnetic field strengths to be expressed in terms of those of the adjacent layer. For transparent films this is given for the r -th layer in the form [29.140]

$$M_r = \begin{pmatrix} \cos \delta_r & i \sin \delta_r / n_r \\ i n_r \sin \delta_r & \cos \delta_r \end{pmatrix}, \quad (29.62)$$

where δ_r is the phase thickness and n_r is the refractive index of the r -th layer. The electric and magnetic field intensities in the adjacent $(r-1)$ -th layer, E_{r-1} and H_{r-1} , are related to those in the r -th layer E_r and H_r by

$$\begin{pmatrix} E_{r-1} \\ H_{r-1} \end{pmatrix} = M_r \begin{pmatrix} E_r \\ H_r \end{pmatrix}. \quad (29.63)$$

Repeated application of (29.63) to a stack of layers enables E_0 and H_0 in the incident medium (n_0) to be expressed in terms of the field intensities in any other layer of the stack. Knowing the n values for each of the various layers of the stack, a *characteristic matrix* (also 2×2) for the entire stack is obtained by repeated matrix multiplication of the individual layer matrices, and allows E_0 and H_0 to be obtained in terms of the transmitted field intensities at the opposite side of the stack. The energy reflectance is then easily obtained. Similar matrices to those in (29.62) may also be defined for the case of absorbing films, although it is then usual to work in terms of the layer admittance η , defined as the ratio of the transverse components of the magnetic to the electric vectors [29.138].

A relatively simple example of an interesting system is a stack of layers of equal optical thickness, but with alternating high (H) and low (L) refractive indices [29.140]. If the optical thicknesses correspond to $\lambda/4$ (a quarter-wave stack), there is a highly reflective region at the wavelength λ , with other highly reflecting regions at 2λ , 3λ , etc. The reflectance value becomes higher the more layers there are in the stack, and approaches unity for a few layers. If the refractive index of the incident medium is n_0 , that of the substrate n_S , and those of the high- and low-index regions n_H and n_L , respectively, the energy reflectance for an even number of layers $2s$ (n_L next to substrate) is given by

$$R_{2s} = \left(\frac{n_S f - n_0}{n_S f + n_0} \right)^2 \quad (29.64)$$

where $f = (n_H/n_L)^{2s}$. For an odd number of layers ($2s+1$), with n_H next to the substrate, the energy reflectance becomes

$$R_{2s+1} = \left(\frac{n_H^2 f - n_0 n_S}{n_H^2 f + n_0 n_S} \right)^2. \quad (29.65)$$

The bandwidth of the highly reflecting region depends on $\sin^{-1}[(n_L - n_H)/(n_L + n_H)]$, and thus is limited by the refractive indices available. Bandwidth can be increased by using a stack with layer thicknesses increasing in geometrical progression.

In addition to antireflection and highly reflecting coatings, various types of optical filters can be designed. Perhaps the simplest is the Fabry–Perot interference filter, which consists of two highly reflecting stacks, separated by a layer of optical thickness $\lambda/2$. This results in a very narrow transmission band at wavelength λ , with a rejection region on either side. Filter designs are usually based on computer evaluations of the optical matrices. These include band-pass, band-stop and edge filters. These are well beyond the scope of the present article and the reader is referred to the literature [29.30, 138, 141, 142].

In this section some of the basic optical properties have been described, with the primary emphasis on non-absorbing films at normal incidence. Several interesting film arrangements have been described, again for non-absorbing films. Implicit to any system there is some absorption, but this can be accounted for if both the optical constants are known. Usually the design process involves the manipulation of several layer matrices, but this is handled effectively by modern computer technology. The optical properties of highly absorbing films are described by *Abelès* [29.143], who also considers inhomogeneous films. Other useful references on the theory and detailed optical properties of various aspects of thin films may be found in the literature: theory and calculations [29.144], antireflection coatings [29.145], filters [29.142, 146, 147] and mirrors [29.148].

29.3.2 Electrical Properties

The electrical properties of thin films are one of their most important physical features, and differ from those of the corresponding bulk material for various reasons. Electrical properties cover a particularly wide field, and in this section only two important examples are covered. The first of these is the lateral conductivity (or corresponding resistivity) of thin metallic films. In metals we are usually concerned only with one type of charge carrier, electrons, and thus variations in conductivity σ due to the presence of both carrier types (such as compensation effects) are not considered. The main cause of deviations from the bulk material conductivity values is curtailment of the bulk electron mean free path λ_0 by the dimensions of the film. Clearly, if the film thickness $d < \lambda_0$, many instances of electron scattering occur at the film surfaces, thus reducing the effective electron mean free path and decreasing the conductivity. Modern theories also take into account electron scattering by individual grain boundaries as well as by the film surfaces. The second example of electrical properties in thin films

is that of high-field conduction in insulators and semiconductors. Electric fields in thin films may approach the dielectric breakdown strength, and under these conditions the usual ohmic behaviour is not followed, but a range of different limiting conduction processes may take place. Conduction in semiconductors at somewhat lower electric fields is not discussed here, but is covered in detail elsewhere in the literature [29.1, 15, 18, 149]. Semiconductor conductivity is a broad subject, underlying the operation of many different electronic devices. We might note that the conductivity of semiconductors in their surface regions (and therefore also in semiconductor thin films) differs from that of the bulk material owing to depletion or accumulation regions at the surfaces, and that the application of a surface electric field forms the basis of several semiconductor devices, such as metal–oxide–semiconductor (MOS) transistors. Furthermore, the mobility of carriers in semiconductors frequently follows similar variations to those of the conductivity in metallic films, for essentially the same regions, in that the surface and/or grain boundary scattering significantly influence the mobility and thus the conductivity.

Electrical conductivity in metallic films

The electrical conductivity in bulk metals σ_0 is given generally by an expression of the form

$$\sigma_0 = \frac{ne^2\lambda_0}{mv}, \quad (29.66)$$

where n is the concentration of free electrons, e is the electronic charge, λ_0 the bulk electron mean free path, m the electron effective mass and v the electron Fermi velocity. This expression results from the free-electron gas theory of metals, derived from the work of Drude, Lorentz and Sommerfeld. The main feature of this result is that the electrical conductivity is proportional to the mean free path, and that if the mean free path is curtailed by any means, then the conductivity will be reduced, all other considerations remaining unaltered. A further consideration is the well-known result termed Matthiessen's rule,

$$\rho_0 = \rho_d + \rho_{ph}(T), \quad (29.67)$$

where the bulk resistivity $\rho_0 (= 1/\sigma_0)$ is given by the sum of contributions from static defects and impurities ρ_d and from phonons $\rho_{ph}(T)$. The latter contribution effectively isolates the main temperature-dependent effect from other scattering mechanisms, unless they too are temperature-dependent. Temperature-dependent

phonon interactions increase with increasing temperature, giving rise to the positive temperature coefficient of resistivity (TCR) in metals. At very low temperatures $\rho_{\text{ph}}(T) \rightarrow 0$, and only the residual resistivity ρ_d remains. For the case of a thin metal film, scattering at the surfaces also comes into play, and Matthiessen's rule is modified by the addition of an extra surface-scattering resistivity term ρ_s to give

$$\rho_f = \rho_d + \rho_{\text{ph}}(T) + \rho_s \quad (29.68)$$

for the thin-film resistivity ρ_f . For very thick films $\rho_s \approx 0$, since the mean free path is not then limited by the surfaces, and the film resistivity is effectively the same as the bulk value. When ρ_s is significant for thinner films, $\rho_f > \rho_0$, and the film conductivity $\sigma_f < \sigma_0$. The above discussions follow naturally from the treatise in Chapt. 2.

Over a century ago, Thomson [29.151] was the first to suggest that observed decreases in the conductivity of thin films as compared to the bulk values are the result of the limitation of the electron mean free path by the specimen size. He made the following assumption in obtaining an expression for the film conductivity: the bulk mean free path in the metal is a constant λ_0 , greater than the film thickness d , and that when electrons collide with the internal film surfaces the scattering is independent of the angle at which they strike, i.e. the scattering is diffuse. Collisions occurring in the film were split into two regions, in the first the mean free path is determined by the surfaces (top and bottom) and in the second the mean free path is not limited by scattering at the surfaces. In terms of the parameter

$$k = \frac{d}{\lambda_0}, \quad (29.69)$$

the ratio of the film thickness to the bulk mean free path, the ratio of the film to bulk conductivity was determined as

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} = \frac{k}{2} \left[\ln\left(\frac{1}{k}\right) + \frac{3}{2} \right]. \quad (29.70)$$

Fuchs [29.150] pointed out that this derivation neglected mean free paths starting at the film surfaces, and also the statistical distribution of the mean free paths about λ_0 in the bulk metal. A corrected version of Thomson's model for very thin films, i.e. $d \ll \lambda_0$ ($k \ll 1$), where effectively all the mean free paths start from one of the surfaces, yields an alternative result [29.152]

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} = k \left[\ln\left(\frac{1}{k}\right) + 1 \right]. \quad (29.71)$$

Although both expressions predict a decrease in conductivity below the bulk value for films, neither gives a sensible prediction for very thick films, i.e. $d \gg \lambda_0$ ($k \gg 1$).

A more useful general model to account for variations of conductivity with film thickness was introduced by Fuchs [29.150], who formulated the problem in terms of the Boltzmann transport equation, which describes the way in which the electron energy changes due to applied fields and collisions. Readers are referred to the original paper, or to one of the reviews mentioned at the end of this section, for details of the mathematical arguments. The final result of this model is

$$\begin{aligned} \frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} = 1 + \frac{3}{4} \left(k - \frac{k^3}{12} \right) B(k) \\ - \frac{3}{8k} (1 - e^{-k}) - \left(\frac{5}{8} + \frac{k}{16} + \frac{k^2}{16} \right) e^{-k}, \end{aligned} \quad (29.72)$$

where $B(k)$ is a representative of the set of exponential integral functions, given by

$$B(x) = -\text{Ei}(-x) = \int_x^\infty \frac{e^{-y}}{y} dy. \quad (29.73)$$

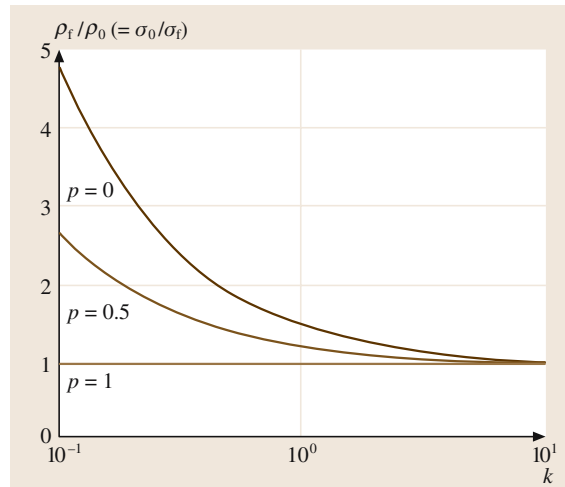


Fig. 29.21 Dependence of the ratio of film to bulk resistivity ρ_f/ρ_0 on the ratio of film thickness to bulk mean free path $k = d/\lambda_0$ predicted by the model of Fuchs [29.150]; p is the specularity parameter, with $p = 0$ representing totally diffuse scattering and $p = 1$ representing totally specular scattering, where for the latter $\rho_f/\rho_0 = 1$. The curve labelled $p = 0.5$ is intermediate between these two values

Figure 29.21 shows the dependence of ρ_f/ρ_0 ($=\sigma_0/\sigma_f$) on k over the range 0.1–10 for the curve labelled $p = 0$ (diffuse scattering). Here it is clear that the resistivity drops to the bulk value for thick films. Various approximations for (29.72) are frequently quoted to aid computation in specific thickness ranges. For thick films ($k \gg 1$)

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx 1 - \frac{3}{8k}. \quad (29.74)$$

This clearly shows that as $k \rightarrow \infty$, $\sigma_f \rightarrow \sigma_0$. For thinner films ($k \ll 1$), Campbell [29.153] quotes

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4} \left[\ln\left(\frac{1}{k}\right) + 0.4228 \right] \quad (29.75)$$

and for very thin films Sondheimer [29.154] gives

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4} \left[\ln\left(\frac{1}{k}\right) \right]. \quad (29.76)$$

Equations (29.70), (29.71), (29.75) and (29.76) give some confidence in the general approach, since they are all of comparable form, with the right-hand sides each consisting of a constant $\times k \times [\ln(1/k) + \text{another constant}]$.

Fuchs also considered the case where the electron scattering at the surfaces was not totally randomised, and that a constant proportion of electrons p are scattered specularly. p is termed the *specularity parameter*, where $p = 0$ for the case of totally diffuse scattering and $p = 1$ for totally specular scattering. Clearly in the latter case the film surfaces do not affect the conductivity, and σ_f reduces to the bulk value. Figure 29.21 also shows the variation of ρ_f/ρ_0 for $p = 0.5$ (intermediate between diffuse and specular scattering) and $p = 1$ (totally specular scattering). Again for the case of some specular scattering there are limiting expressions for thick films ($k \gg 1$) and thin films ($k \ll 1$), given respectively by

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx 1 - \frac{3(1-p)}{8k} \quad (29.77)$$

and

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4} (1+2p) \left[\ln\left(\frac{1}{k}\right) + 0.4228 \right], \quad (29.78)$$

where p is small in (29.78). For totally diffuse scattering, where $p = 0$, these two expressions reduce to the previous results of (29.74) and (29.75). An alternative expression for thin films is also given by Sondheimer [29.154],

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4} \left(\frac{1+p}{1-p} \right) \ln\left(\frac{1}{k}\right). \quad (29.79)$$

This reduces to (29.76) for $p = 0$, and for small k and p is consistent with (29.78).

Both Lucas [29.155] and Juretschke [29.156, 157] have extended the specularity analysis by assuming a different specularity parameter for each side of the film, i. e. p and q . This is of course quite reasonable, in that the specularity parameter of the lower (substrate) surface is likely to be heavily influenced by the substrate surface. The thick- ($k \gg 1$) and thin-film ($k \ll 1$) approximations in this case are then

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx 1 - \frac{3}{8k} \left[1 - \left(\frac{p+q}{2} \right) \right] \quad (29.80)$$

and

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} \approx \frac{3k}{4} \frac{(1+p)(1+q)}{(1-pq)} \ln\left(\frac{1}{k}\right), \quad (29.81)$$

respectively. For identical surfaces with $p = q$, these expressions reduce to (29.77) and (29.79) respectively. Tabulated values of ρ_f/ρ_0 are given by Chopra [29.18] as functions of k for various combinations of p and q . The whole topic of the size-dependent electrical conductivity in thin metallic films and wires is thoroughly reviewed by Larson [29.158], but omitting the important work on grain-boundary scattering of Mayadas and Shatzkes [29.159]. This is, however, covered in a more recent book reviewing size effects in thin films generally [29.160], which also includes useful tabulated conductivity data.

As mentioned above, an important development in size-effect models came about with the inclusion of grain-boundary scattering [29.159]. In a bulk sample, the individual grains are normally considerably larger than the bulk mean free path λ_0 , and do not therefore affect the conductivity. However, in thin films, the grain sizes and orientations depend on the preparation method and conditions. Typical grain sizes are of the same order as λ_0 and therefore scattering at their boundaries will influence the conductivity. Mathematically, the model takes account of the fact that grains tend to grow in a columnar structure, with the axes normal to the film plane. Single grains usually extend between both film surfaces, and thus their upper and lower boundaries do not affect the conductivity. Mayadas and Shatzkes argue that only grain boundaries normal to the film plane influence the conductivity. The grain boundaries are represented by δ -function potentials, and are used in the model in the solution of the Boltzmann transport equation, subject to suitable boundary conditions. The model is developed in terms of a dimensionless parameter α , given by

$$\alpha = \frac{\lambda_0}{D} \frac{R}{1-R}, \quad (29.82)$$

where, as before λ_0 is the bulk mean free path, D is the mean grain dimension, and R is the grain-boundary electron reflection coefficient. The ratio of the grain-boundary conductivity σ_g (and the corresponding resistivity ρ_g) to the bulk values are then given by

$$\frac{\sigma_g}{\sigma_0} = \frac{\rho_0}{\rho_g} = 3 \left[\frac{1}{3} - \frac{1}{2}\alpha + \alpha^2 - \alpha^3 \ln \left(1 + \frac{1}{\alpha} \right) \right]. \quad (29.83)$$

This is easy to evaluate, and the dependence of ρ_g/ρ_0 ($= \sigma_0/\sigma_g$) on α is shown in Fig. 29.22. For small values of α (i. e. large grain size D , small bulk mean free path λ_0 , or small reflection coefficient R) the grain boundaries do not affect the overall conductivity, which remains at the bulk value. However, for large α the grain-boundary resistivity increases (conductivity decreases). For very large ($\alpha \gg 1$) and very small ($\alpha \ll 1$) values of α , (29.83) has the limiting approximations

$$\frac{\sigma_0}{\sigma_g} = \frac{\rho_g}{\rho_0} \approx \frac{4}{3}\alpha \quad (29.84)$$

and

$$\frac{\sigma_0}{\sigma_g} = \frac{\rho_g}{\rho_0} \approx 1 + \frac{3}{2}\alpha, \quad (29.85)$$

respectively. Note that following the original work [29.159] the latter two expressions are quoted as the ratio ρ_g/ρ_0 ($= \sigma_0/\sigma_g$), i. e. the reciprocal of the ratio in (29.83).

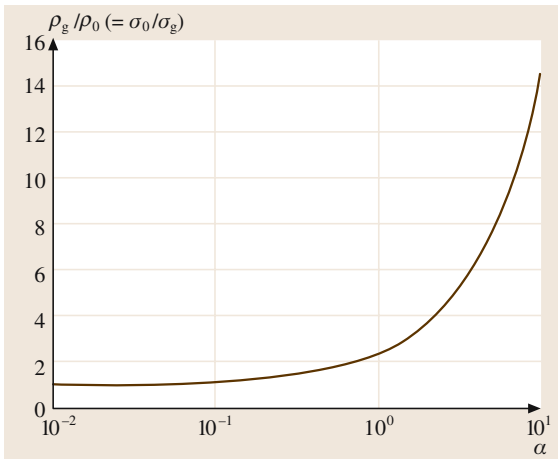


Fig. 29.22 Dependence of the ratio of grain-boundary to bulk resistivity ρ_g/ρ_0 on the parameter α from the model of Mayadas and Shatzkes [29.159]. The data are calculated from (29.83)

Mayadas and Shatzkes [29.159] also considered the case of grain-boundary scattering in conjunction with scattering at the film surfaces, and followed the Fuchs [29.150] formulation to arrive at an expression for the overall film resistivity ρ_f . Unfortunately, the resulting expression could not be evaluated analytically, requiring numerical solution. Detailed comparisons between the basic Fuchs model with $p = 0$ and $p = 0.5$, and their model using the same values of p and including grain-boundary scattering with $\alpha = 1$, led these workers to interpret their results as showing an enhancement of the intrinsic resistivity ρ_g in polycrystalline samples in which grain size is constant and independent of film thickness. However, discrepancies between the Fuchs and Mayadas–Shatzkes models become most apparent for very large α (very small grain size, D). For this situation the deviation corresponds to the additional resistivity arising from grain boundaries rather than from surface scattering [29.161, 162]. Useful tabulated data of resistivity ratios in samples with both grain-boundary and surface scattering are given elsewhere in the literature [29.160, 163].

Finally, we briefly discuss the thickness variations of the temperature coefficient of resistivity. This is a measure of the response of the resistivity to changes in temperature, and is usually denoted by the symbol α . To avoid confusion with the Mayadas–Shatzkes parameter, here we denote it by α_T . The temperature coefficient of resistivity is given for the bulk (α_{T0}) and for thin films (α_{Tf}) by

$$\alpha_{T0} = \frac{1}{\rho_0} \frac{d\rho_0}{dT}; \quad \alpha_{Tf} = \frac{1}{\rho_f} \frac{d\rho_f}{dT} \quad (29.86)$$

where ρ_0 and ρ_f are the bulk and thin-film resistivities, as given in (29.67) and (29.68). From the above relationships, and Matthiesen's rule, it is possible to derive expressions for α_{Tf}/α_{T0} in a similar manner as ρ_f/ρ_0 . Referring to (29.68), ρ_d is essentially temperature-independent, while the temperature dependence of ρ_s depends on the film thickness. For thicker films ($k \gg 1$), ρ_s is also temperature-independent, and thus

$$\frac{d\rho_f}{dT} = \frac{d\rho_{ph}}{dT} \quad (29.87)$$

Similarly, from (29.67) for the bulk material

$$\frac{d\rho_0}{dT} = \frac{d\rho_{ph}}{dT} \quad (29.88)$$

and therefore both the differentials in (29.86) are equal, and we obtain

$$\alpha_{T0}\rho_0 = \alpha_{Tf}\rho_f. \quad (29.89)$$

This useful result implies that

$$\frac{\sigma_f}{\sigma_0} = \frac{\rho_0}{\rho_f} = \frac{\alpha_{Tf}}{\alpha_{T0}} \quad (29.90)$$

and that for the case of thicker films the ratio of the film-to-bulk temperature coefficient of resistivity is equal to the ratio of film-to-bulk conductivity. For the case of thinner films ($k \ll 1$), the situation is complicated by the fact that ρ_s is also temperature-dependent, because the mean free path decreases with increasing temperature and becomes significant for small k . Various calculated data for α_{Tf} are given elsewhere in the literature [29.6, 18, 153, 158, 160], to which the reader is referred for a fuller discussion. Discussions concerning the applicability of the various models described above to several types of metallic film are also given in some of these references, and a wider coverage including additional contributions to the size effect models, which cannot be included in the present work, is given in the book by *Tellier and Tosser* [29.160].

High-field conduction in insulating and semiconducting thin films

Here electrical conductivity is considered for the case where a voltage is applied across an insulating or semiconducting thin film prepared in a sandwich configuration between metallic electrodes. This situation is in direct contrast to that described in the previous section, in that the intrinsic film conductivity is significantly lower, and the electric field applied can be extremely high, even when only low voltages are applied. The latter point is clearly illustrated by the fact that in a film of thickness $1 \mu\text{m}$ subjected to a voltage of only 10 V, the mean electric field established across the film is 10^7 V/m , approaching the dielectric breakdown strength of many materials. Several high-field conduction mechanisms that may in principle occur in thin dielectric films have been thoroughly reviewed in the literature [29.164]. These have also been observed in relatively wide-band-gap semiconducting films. These processes are discussed later in this section, together with the appropriate current density–voltage expressions which aid in their identification and analysis. First, however, it is useful to discuss the types of electrical contact which may be applied to dielectric films, since the dominant conduction process is frequently determined by the type of electrical contact and its interfacial properties.

Simmons [29.164] has demonstrated that the type of electrode used significantly influences the conduction processes which may be observed; the various conduction mechanisms were classified as either *bulk-limited*

or *electrode-limited*. For the former the charge carriers are generated in the bulk of the material and the electrodes merely serve to apply a potential which generates a drift current. In the latter, at the interface between the electrode and the insulator or semiconductor, a potential barrier to the flow of charge is established, which limits the current. Contacts themselves are said to be either *ohmic* or *blocking*, with an intermediate *neutral* contact. Blocking contacts are conventionally known as Schottky barriers. In principle, the type of contact formed depends on the relative work functions of the insulator or semiconductor ψ_i and that of the metal contact ψ_m . Excellent energy-band diagrams of the various types of contact are given in the literature [29.164], and are reproduced elsewhere in descriptions of contacts to cadmium compound films [29.38] and to nanostructures [29.11]. In equilibrium the vacuum and the Fermi levels are continuous across the interface, and with no applied voltage the Fermi level also must be flat, in order that no current flows. Away from the interface in the bulk of the material, the energy difference between the vacuum and Fermi levels will be equal to the work function ψ_i , and it can easily be seen that the potential barrier height at the interface φ_0 is given by

$$\varphi_0 = \psi_m - \chi, \quad (29.91)$$

where χ is the film electron affinity and ψ_m is the metal work function. However, it is the *type* of contact rather than its precise barrier height which determines the type of conductivity observed. We consider primarily electrons as the charge carriers, as it has been argued that for insulators conduction by holes may normally be ignored, owing to relatively low mobilities and to immobilisation by trapping effects. Similar arguments are also applicable in the case of wide-band-gap semiconductors, and thus for both insulators and some semiconductors n-type conductivity is usual. For p-type conductivity, the conditions concerning the relative values of the work functions are reversed [29.165].

Ohmic contacts occur for electron injection when $\psi_m < \psi_i$. Electrons are injected from the electrode into the insulator or semiconductor in order to comply with thermal equilibrium requirements. They are subsequently located within the insulator or semiconductor conduction band and penetrate a distance λ_0 below the interface, forming a region of negative space charge which is termed an *accumulation region*. Upwards band bending of the conduction band occurs in the accumulation region, beyond which the dielectric is shielded from the electrode. The accumulation region acts as a charge reservoir, and can supply electrons to the

material as required by the bias conditions. It has been pointed out by *Simmons* [29.164] that the conductivity is limited by the rate at which electrons can flow into the bulk of the material, and is therefore bulk-limited. Conversely, blocking contacts or Schottky barriers occur when $\psi_m > \psi_i$. In this case the initial flow of charge is reversed, with electrons flowing from the insulator or semiconductor to the metal electrode. In contrast to the case for ohmic contacts, this results in a positive space-charge region near the interface in the insulator or semiconductor, which, owing to the absence of electrons, is termed a *depletion region*. An equal and opposite negative charge is induced on the electrode, and the interaction between the two charges establishes a local electric field near the interface. In this case the conduction band in the insulator or semiconductor bends downwards, and the free-electron concentration in the interfacial region is much lower than in the bulk of the film material. The rate of flow of electrons is thus limited by the rate at which they can flow over the interfacial barrier, and the conductivity is electrode-limited.

Although, in principle, the type of contact established depends on the relative values of ψ_m and ψ_i , in reality such considerations are usually inadequate in predicting the type of contact formed, because of the existence of *surface states*. These are due either to the effects of unsaturated bonds and impurities at the interface [29.166] or to the departure from periodicity in the structure at the interface. It is very difficult to accurately predict a priori the presence and concentration of surface states. If a significant concentration of surface states does exist, (29.91) is inapplicable and the type of contact and the conduction process are determined almost entirely by their presence.

In the remainder of this section the bulk-limited conduction processes are first considered, where the conductivity does not depend on the barrier height at the interface. Following this the electrode-limited processes are considered, where the barrier height becomes a significant feature in the current density–voltage (J – V) equations.

A particularly important conduction process in insulating and semiconducting thin films is that of space-charge-limited conductivity (SCLC). For SCLC to occur there is a requirement that the injecting electrode is an ohmic contact. Thus there is a charge reservoir available in the accumulation region, and electrons do not need to be excited over a potential barrier. At low voltages the thermally generated carrier concentration exceeds the injected concentration, and the current density J is proportional to the applied voltage V , following Ohm's

law,

$$J = n_0 e \mu \frac{V}{d}, \quad (29.92)$$

where n_0 is the thermally generated carrier concentration (electrons), e is the electronic charge, μ the mobility and d the film thickness. When the injected electron concentration exceeds that of the thermally generated concentration the SCLC current becomes dominant. SCLC current density–voltage relationships are derived by taking into account both the drift and diffusion currents, and solving Poisson's equation. This results in an expression for the electric field as a function of distance from the interface. Charge-carrier traps normally exist within the insulating or semiconducting film, and these are effective in mopping up or immobilising most of the injected electrons. For shallow traps located at a discrete energy E_t below the conduction band edge the SCLC current density is given by [29.164]

$$J = \frac{9}{8} \varepsilon_r \varepsilon_0 \theta \mu \frac{V^2}{d^3}, \quad (29.93)$$

where ε_r is the relative permittivity of the material, ε_0 the permittivity of free space and θ is the ratio of free to trapped charge. This expression predicts that $J \propto V^2$ and $J \propto d^{-3}$, and the search for these relationships is an important tool in identifying SCLC dominated by a discrete trap level. For perfect materials which do not contain traps, θ is replaced by unity in (29.93). The ratio of free to trapped charge is given by [29.167]

$$\theta = \frac{N_c}{N_{t(s)}} \exp\left(-\frac{E_t}{k_B T}\right), \quad (29.94)$$

where N_c is the effective density of states at the conduction band edge, $N_{t(s)}$ is the shallow trap concentration located at the discrete energy level, k_B is Boltzmann's constant and T is the absolute temperature. A transition between ohmic conductivity and SCLC takes place when the injected carrier concentration first exceeds the thermally generated carrier concentration. An expression for the transition voltage V_t at which this occurs is obtained by simultaneously solving (29.92) and (29.93) to yield

$$V_t = \frac{8en_0d^2}{9\theta\varepsilon_r\varepsilon_0}. \quad (29.95)$$

SCLC of this type, showing square-law J – V behaviour, has been observed in many types of thin films. In SbS_3 dielectric films, with bismuth oxide ohmic contacts, the current was at first ohmic, before a transition to SCLC [29.168]. Current density was also found to be

proportional to $1/d^3$ at constant voltage, and the transition voltage V_t showed the characteristic linear variation with d^2 . It has also been observed in CdS films [29.40, 169, 170], particularly with In injecting electrodes, and a discrete-trap concentration of $2.5 \times 10^{20} \text{ m}^{-3}$ was deduced [29.40]. In CdTe, a square-law SCLC behaviour was recorded in electro-deposited Ni–CdTe–Au structures [29.171]. The work also showed a systematic increase in current density with temperature over the range 273–338 K, together with the expected $J \propto 1/d^3$ behaviour. Dharmadhikari [29.172] observed similar behaviour in evaporated CdTe films in the lower voltage range, although power-law behaviour became apparent at higher voltages. CdSe films have also been reported as showing square-law SCLC [29.173, 174]. In organic films, the phthalocyanines are well known for exhibiting SCLC behaviour [29.111]. Square-law behaviour with discrete trap levels has been seen in several of these materials, with typical activation energies of 0.77 eV [29.175] and 0.62 eV [29.176] in copper phthalocyanine, 0.42 eV in lead phthalocyanine [29.177], 0.58 eV in nickel phthalocyanine [29.178] and 0.45 eV and 0.69 eV in cobalt phthalocyanine [29.179]. These values appear to vary according to the central metal atomic species in the molecule, and also with the oxygen content. In general these materials are p-type, having a trap concentration for discrete levels in the range $5 \times 10^{16} - 7.1 \times 10^{25} \text{ m}^{-3}$.

For amorphous and polycrystalline materials, where a single shallow trap level is unlikely to be present, several poorly defined trap levels can lead to an overall exponential distribution of traps $N(E)$ [29.167] given by

$$N(E) = N_0 \exp\left(-\frac{E}{k_B T_t}\right), \quad (29.96)$$

where $N(E)$ is the trap concentration per unit energy range at an energy E below the conduction band edge, N_0 is the value of $N(E)$ at the conduction band edge and $T_t > T$ is a temperature parameter which characterises the distribution. Furthermore, it can be shown that the total concentration of traps comprising the distribution $N_{t(e)}$ is [29.181]

$$N_{t(e)} = N_0 k_B T_t. \quad (29.97)$$

Lampert [29.182] obtained an expression for the current density by assuming an exponential trap distribution as described by (29.96) to yield

$$J = e\mu N_c \left(\frac{\varepsilon_r \varepsilon_0}{e N_0 k_B T_t}\right)^l \frac{V^{l+1}}{d^{2l+1}}. \quad (29.98)$$

This expression predicts a power-law dependence of J on V , with the exponent $n = l + 1$, where l represents the ratio T_t/T . J is also expected to be proportional to $d^{-(2l+1)}$. Similarly to the case for a single shallow trap level, there is also a transition voltage between ohmic conduction and SCLC dominated by an exponential trap distribution. The transition voltage in this case is given by [29.183]

$$V_t = \left(\frac{n_0}{N_c}\right)^{1/l} \frac{e N_0 k_B T_t d^2}{\varepsilon_r \varepsilon_0}. \quad (29.99)$$

Many instances of power-law SCLC are also reported in the literature. Zuleeg [29.169] observed a cubic power-law dependence at higher voltages in his CdS films, which suggested that an exponential trap distribution was present. Dharmadhikari [29.172] appears to have been the first to observe a power-law J – V dependence in CdTe films at higher voltage levels, in addition to the square-law dependence at lower voltages. Very high exponents were recorded, although at the time this was not analysed using the exponential trap distribution expressions given above. This connection was made in a later series of measurements [29.180], where the intermediate square-law behaviour was absent. This

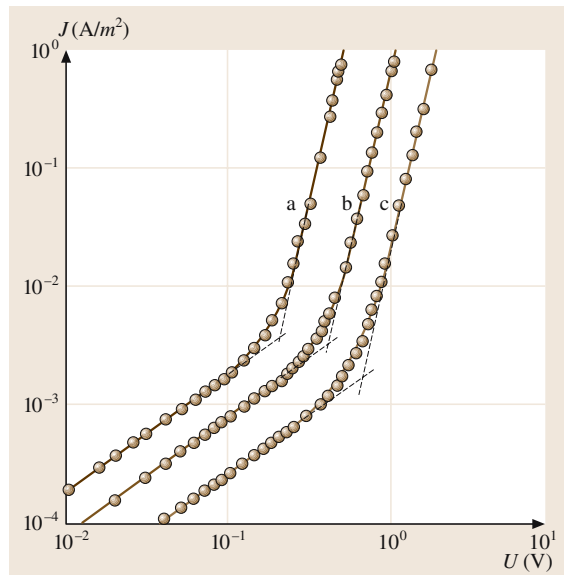


Fig. 29.23a–c Room-temperature current density–voltage characteristics for evaporated Al–CdTe–Al sandwich structures of thickness 0.31 μm (a), 0.45 μm (b) and 0.56 μm (c), showing ohmic conduction and SCLC. U is the applied voltage. After [29.180] with permission from Wiley–VCH

J - V dependence is shown in Fig. 29.23 [29.180] for three different film thicknesses, and is typical of many results on various different materials showing an exponential trap distribution. Similar exponential trap distributions have also been identified in some CdSe films [29.184–186], notwithstanding the earlier results which indicated discrete trap levels [29.173, 174]. The differences between these two classes of results appear to depend on the different deposition conditions and processing used. Exponential trap distributions are extremely prevalent in phthalocyanine films, and were originally investigated by *Sussman* [29.187] for copper phthalocyanine films with ohmic Au electrodes. The material was shown to be p-type, and exhibited exponent values n in the J - V characteristics in the range 2.6–4.0. The predicted thickness dependence for this type of conductivity was also observed. Many other phthalocyanine films have subsequently shown this type of conductivity, and data on these are tabulated and compared in the literature [29.111]. Total trap concentrations for the exponential distributions are in the range 6×10^{20} – $9.3 \times 10^{26} \text{ m}^{-3}$.

Using the above Eqs. (29.96–29.98) as a starting point, it was shown by the present author [29.188] that for SCLC dominated by an exponential trap distribution, measurements of J as a function of temperature at constant applied voltage in the SCLC region are sufficient to determine both the mobility and the trap concentration. If the data are plotted in the form $\log_{10} J$ against $1/T$ the curves should be linear, and when extrapolated to negative values of $1/T$, should all intersect at a common point irrespective of the applied voltage. The coordinates of this point are given by

$$\log_{10} J = \log_{10} \left(\frac{e^2 \mu d N_c N_{t(e)}}{\varepsilon_r \varepsilon_0} \right); \quad \frac{1}{T} = -\frac{1}{T_t}. \quad (29.100)$$

Other useful results enabling data to be extracted from these plots are that for the gradient, given by

$$\frac{d(\log_{10} J)}{d(1/T)} = T_t \log_{10} \left(\frac{\varepsilon_r \varepsilon_0 V}{e d^2 N_{t(e)}} \right) \quad (29.101)$$

and for the intercept $\log_{10} J_0$ on the $\log_{10} J$ axis, which is

$$\log_{10} J_0 = \log_{10} \left(\frac{e \mu N_c V}{d} \right) \quad (29.102)$$

This set of equations has been used to determine mobility and trap concentration in several different materials showing this form of SCLC, including copper phthalocyanine, where these quantities were measured as

functions of the ratio of evaporation background pressure to deposition rate [29.189]. In another example the method was used to determine these quantities in semiconducting CdTe films [29.190].

We might mention that *Rose* [29.167] also explored the possibility of a uniform trap distribution, in addition to discrete energy levels and the exponential distribution. In this case $J \propto V \exp(tV)$, where t is a temperature-dependent constant, but this is observed far less frequently than the square-law and power-law behaviour described above. An example is in the case of electrodeposited p-type CdTe films [29.191], where *Rose's* expression was followed and the trap concentration per unit energy range was $10^{21} \text{ eV}^{-1} \text{ m}^{-3}$. It was also observed in early work on copper phthalocyanine films with ohmic Au electrodes [29.192]. The discussion above for all types of SCLC considers carrier injection of a single type only. Double injection, when one electrode is ohmic for electrons and the other ohmic for holes, is covered in detail in the text by *Lampert* and *Mark* [29.193] on current injection in solids, and is also discussed by *Lamb* [29.166].

The second bulk-limited conduction process to be discussed is the Poole–Frenkel effect, which is essentially the field-assisted lowering of the Coulombic potential barrier ϕ between electrons located at impurity sites and the edge of the conduction band. Electrons in such centres are unable to contribute to the conductivity until they overcome the potential barrier ϕ and are promoted into the conduction band. For an applied electric field F the potential is reduced by an amount eFx where e is the electronic charge and x is the distance from the centre. For high electric fields the lowering of the potential can be significant. Such a variation is illustrated in the literature, usually for the case of a donor level located an energy E_d below the bottom of the conduction band [29.164]. The potential energy of the electron in the Coulombic field is $-e^2/4\pi\varepsilon_r\varepsilon_0x$, and there is also another contribution to the potential energy, $-eFx$, resulting from the applied field F . The combined effect of both these contributions to the potential energy is that there is a maximum in the potential energy for emission at a distance x_m from the centre. The effective potential barrier for Poole–Frenkel emission is then lowered by an amount $\Delta\phi_{\text{PF}}$, which depends on the electric field according to the relationship

$$\Delta\phi_{\text{PF}} = \beta_{\text{PF}} F^{1/2}, \quad (29.103)$$

where

$$\beta_{\text{PF}} = \left(\frac{e^3}{\pi\varepsilon_r\varepsilon_0} \right)^{1/2} \quad (29.104)$$

is referred to as the Poole–Frenkel field-lowering coefficient. The conductivity, and consequently the current density, depend exponentially on the potential barrier for excitation of electrons into the conduction band, and since this is lowered by $\Delta\varphi_{\text{PF}} = \beta_{\text{PF}} F^{1/2}$ the current density is increased by an exponential factor. The current density is given by

$$J = J_0 \exp\left(\frac{\beta_{\text{PF}} F^{1/2}}{kT}\right), \quad (29.105)$$

where $J_0 = \sigma_0 F$ is the low-field current density. Although J_0 in this expression depends linearly on F , it is frequently ignored in comparison with the exponential term, which increases far more rapidly with F . Equation (29.105) is usually applicable for thin-film insulators and semiconductors in the presence of traps. If, rarely, traps are absent and the low-field conductivity σ_0 shows an intrinsic dependence proportional to the factor $\exp(-E_g/2kT)$, where E_g is the energy gap, then a factor of 2 appears in the denominator of the exponential term in (29.105) [29.164]. Since $F = V/d$, the above expression can be written

$$J = J_0 \exp\left(\frac{\beta_{\text{PF}} V^{1/2}}{kTd^{1/2}}\right). \quad (29.106)$$

Therefore a plot of $\log J$ against $V^{1/2}$ should show a linear relationship in the case of Poole–Frenkel conductivity, from which a value of the field-lowering coefficient can be deduced and compared with the theoretical value of (29.104). If it is necessary for the variation of the pre-exponential factor to also be taken into account, then a linear plot of $\log(J/V)$ against $V^{1/2}$ should be obtained.

Modifications to (29.105) have been proposed to account for the somewhat different current density dependence in the presence of various combinations of donors and traps. *Simmons* [29.164] mentioned the case where the material contains donor levels below the Fermi level as well as shallow neutral traps, in which the coefficient of $F^{1/2}/kT$ is one half of the theoretical value given by (29.104). For the case of a nonuniform electric field, having a maximum value of $\alpha^2 F$, where F is the mean field, the following expression has been proposed [29.195]

$$J = \frac{2J_0 kT}{\alpha \beta_{\text{PF}} F^{1/2}} \exp\left(\frac{\alpha \beta_{\text{PF}} F^{1/2}}{k_B T}\right). \quad (29.107)$$

Although this expression was originally proposed for the case of CdTe films, it has also been adopted for the interpretation of results in other materials.

The Poole–Frenkel effect has been observed in many types of thin film, including insulating oxides, semiconductors and organic materials. The main indicator of Poole–Frenkel conductivity is a linear dependence of $\log J$ on $V^{-1/2}$, although it is also seen in the case of the Schottky effect, as will be discussed later. Care must be taken in distinguishing between the two processes. In SiO films, good agreement with (29.106) was obtained [29.196], but the value of β derived exceeded the theoretical Poole–Frenkel value. This type of discrepancy had previously been discussed in the literature [29.197], and it was later proved that in this model the J – V characteristics were symmetric over five orders of magnitude of applied voltage, irrespective of electrode material, which ruled out the Schottky mechanism [29.198]. In another important insulator, silicon nitride, *Sze* [29.199] observed Poole–Frenkel behaviour in metal–insulator–semiconductor (MIS) diodes fabricated on a degenerate Si substrate, over a range of nearly four decades in current. It was found that the current density versus electric field characteristics were independent of film thickness, sample area, electrode materials and polarity, suggesting that the

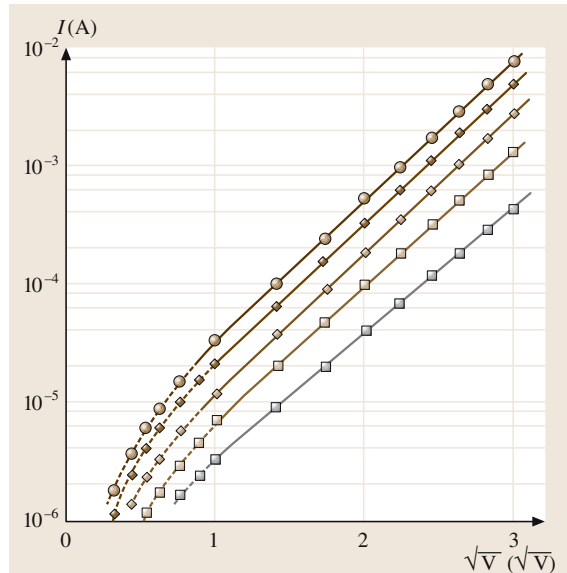


Fig. 29.24 Linear dependence of the logarithm of current on the square root of the applied voltage in an evaporated Al–CdTe–Al sandwich structure of thickness 722.5 nm at 0 °C (bottom curve), 30 °C, 55 °C, 80 °C and 108 °C (top curve). The behaviour was identified with a form of the Poole–Frenkel effect. After [29.194] with permission from Elsevier Science

conductivity was bulk-limited rather than electrode-limited, and therefore eliminating the Schottky effect. Poole–Frenkel conductivity has also been observed in several semiconducting cadmium compounds, such as RF-sputtered CdS films [29.200, 201] and more widely in CdTe films [29.194, 195, 202]. In Al–CdTe–Al samples, Poole–Frenkel-type conductivity was consistently observed over the temperature range 0–108 °C. These results are typical of Poole–Frenkel behaviour, and are illustrated in Fig. 29.24 [29.194]. However, the value of the experimental field-lowering coefficient increased from $4.95 \times 10^{-5} \text{ eV m}^{1/2} \text{ V}^{-1/2}$ at 0 °C to $7.7 \times 10^{-5} \text{ eV m}^{1/2} \text{ V}^{-1/2}$ at 108 °C, and was about twice the theoretical β_{PF} for CdTe. The Schottky effect was eliminated, as the results did not appear to depend on the electrode workfunction when the Al electrodes were replaced with either In or Ag. These workers adopted the suggestion of *Jonscher* and *Ansari* [29.198] that electrons can hop between sites as the result of thermal activation; on this assumption good agreement was found with the predicted β_{PF} , with an activation energy between hopping sites of 0.15 eV. Very similar results to these were also observed subsequently [29.195], and again the measured value of β exceeded the theoretical Poole–Frenkel value. This coefficient was also calculated from the gradients of the $\ln J-V^{1/2}$ characteristics as a function of temperature, and consistent values were obtained. It was suggested that the similarity between the results of this work and the earlier work [29.194] was related to individual centres experiencing varying electric fields, i.e. the electric field was nonuniform. Equation (29.107) was proposed to account for this effect, and has been successful in accounting for the enhanced coefficients sometimes obtained. Poole–Frenkel conductivity was also identified in p-type CdTe, which also shows a change in conductivity type from SCLC to Poole–Frenkel conductivity [29.203]. This conduction mechanism has also been observed in other cadmium compounds, namely CdSe [29.186] and Cd₃As₂ [29.204]. In phthalocyanine films several instances of this effect have been reported. In oxygen-doped copper phthalocyanine films with a Au and an Al electrode, slightly enhanced values of β were obtained and attributed to a nonuniform electric field, whereas in annealed films the field appeared to be uniform [29.205]. Agreement with the basic model was also found for samples with a Au and a Pb electrode [29.175]. Lead phthalocyanine films with Au and Al electrodes exhibited two different values of β [29.177] consistent with Poole–Frenkel conductivity at higher fields and the Schottky effect at lower fields. However other work-

ers had not considered that Poole–Frenkel conductivity was apparent in the same system [29.206], preferring an interpretation based on two different Schottky barrier widths. Both Schottky and Poole–Frenkel behaviour were observed in oxygen-containing lead phthalocyanine films with two Au electrodes [29.207].

Hopping is a third bulk-limited conduction process which is observed in thin films. This type of conductivity is observed particularly in noncrystalline materials, and is thoroughly discussed in various texts concerned with this topic [29.121, 208]. In this type of material the lack of long-range order results in a phenomenon known as localisation, where the energy levels do not merge into a continuum, particularly in the band-tail regions at the edge of the energy bands. Electrons are transported through the material in a series of jumps or *hops* from one localised energy level to another. The hopping process can occur between the localised energy levels even when only small amounts of thermal energy are available, because the localised levels are normally very closely spaced in energy. For this reason, the process can take place at very low temperatures when other processes are energetically impossible. *Mott* and *Davis* [29.121] argued that the conductivity σ exhibits different behaviour in different temperature regions. At higher temperatures thermal excitation of carriers to the band edges is possible and extended-state or free-band conductivity can take place, while at lower temperatures, where less thermal energy is available, hopping may occur. Different varieties of hopping may take place, and these are distinguished by the length of the hop. Nearest-neighbour hopping is self-explanatory, while in variable-range hopping the hops are on average further than to the nearest neighbour, but the energy difference between the states involved is lower. In the case of variable-range hopping the conductivity has been shown to follow a relationship of the form [29.208]

$$\sigma = \sigma_0 \exp\left(-\frac{A}{T}\right)^{1/4} \quad (29.108)$$

where σ_0 and A are constants. This relationship is known as the Mott $T^{-1/4}$ law, since a plot of $\log \sigma$ against $T^{-1/4}$ shows a linear characteristic with negative slope.

Hopping conduction has been observed in many materials, and is extensively reviewed in the literature [29.121, 208, 209]. The $T^{-1/4}$ law is observed in amorphous Ge [29.210], with the results of *Walley* and *Jonscher* [29.211] also showing this behaviour [29.121]. There is also an example of its appearance in amorphous carbon films [29.212]. AC hopping effects have been observed in several cadmium chalcogenides, and this

work is reviewed elsewhere [29.38]. DC hopping was identified in CdTe films with Al electrodes [29.213] following earlier work on AC conductivity [29.172]. The results suggested a hopping process at temperatures below about 175 K, with a hopping centre concentration of about $9 \times 10^{12} \text{ m}^{-3}$ and a hopping mobility of $2.16 \times 10^{-5} \text{ m}^2/\text{Vs}$, the latter considerably lower than the free-band mobility applicable at higher temperatures. The extent of the electron wavefunction was estimated to be 0.69 nm, corresponding to the lattice parameter in CdTe, and supporting the existence of a nearest-neighbour hopping process. Hopping at lower fields was also identified in CdSe films [29.214] and associated with an impurity conduction process proposed previously [29.215]. Variable-range hopping following the $T^{-1/4}$ law has been observed in iron phthalocyanine and in triclinic lead phthalocyanine films [29.216, 217], although the temperature range was restricted. In the work on iron phthalocyanine the law was followed over the range 140–220 K, with particularly good agreement in the range 156–175 K. In the lead phthalocyanine films similar very good linear behaviour was observed in the temperature range of approximately 220–260 K. Hopping at low temperatures has also been indicated from AC measurements in various phthalocyanine systems [29.218–221].

The first electrode-limited conduction process to be considered is that of tunnelling, which is a quantum-mechanical effect, in which the electron wavefunction is attenuated only moderately by a *thin* barrier, resulting in there being a finite probability of its existence on the opposite side of the barrier. Tunnelling directly from the Fermi level of one electrode to the conduction band of the other is normally possible only for very thin films of thickness less than about 10 nm when subjected to a high electric field. Simmons [29.222] investigated the effects of tunnelling between similar electrodes separated by a thin insulating film of thickness s , where the barrier presented by the film was of arbitrary shape. The Wentzel–Kramers–Brillouin (WKB) approximation was used to predict the probability of an electron penetrating the barrier. Different approximations for the current density were found, depending on the relative values of the applied voltage V and φ_0/e , where φ_0 is the barrier height at the electrode interface and e is the electronic charge. These approximations are too unwieldy to be included here, however for the case of $V > \varphi_0/e$ it is only necessary for electrons tunnelling from the Fermi level of one electrode to penetrate a distance $\Delta s < s$ to reach unoccupied levels in the second electrode. This worker also suggested [29.223] that for $V > \varphi_0/e$ a modified

Fowler–Nordheim [29.224] expression may be applicable for tunnelling through an interfacial region. In this case, the electric field at the barrier is sufficiently high to reduce the barrier width, measured at the Fermi level, to about 5 nm. Under these circumstances the current density is related to the voltage and barrier thickness according to the expression

$$J = \frac{e^3 V^2}{8\pi h \varphi_0 d_t^2} \exp\left(-\frac{8\pi(2m)^{1/2} \varphi_0^{3/2} d_t}{3ehV}\right), \quad (29.109)$$

where h is Planck's constant, φ_0 and d_t are the barrier height and effective thickness, respectively, of the tunnelling barrier and m is the free electron mass. The tunnelling current is largely controlled by the barrier height φ_0 , which appears both in the pre-exponential and in the exponential terms. Fowler–Nordheim tunnelling is indicated by a linear dependence of $\log J/V^2$ on $1/V$. A further interesting point for tunnelling in general is that there is very little temperature dependence in the current density, a feature allowing it to be distinguished from other temperature-dependent mechanisms, such as the Schottky effect. There is in fact a very slight quadratic dependence on temperature [29.164], but this is normally negligible within the limitations of experimental measurement.

Early measurements on tunnelling were made by Fisher and Giaever [29.225] on Al_2O_3 produced by oxidation of Al. There was good agreement with an early tunnelling model [29.226], providing the electron effective mass in the insulator was about 1/9 of the free electron mass. Very thin films of Al_2O_3 and BeO showed ohmic and faster-than-exponential dependencies of J on V at lower and higher voltages respectively [29.227]. These results were correlated with direct metal-to-metal electrode tunnelling at low voltages, and from the Al electrode to the insulator conduction band at higher voltages. The model of Simmons [29.164] was tested by measurements on thermally grown Al_2O_3 films on Al, which used several different types of evaporated counter electrodes [29.228]. An extended tunnelling model [29.229], which is applicable to samples with dissimilar electrodes, was followed over nine decades of current. Measurements on thermally grown SiO_2 films, of thickness 65–500 nm with Al or Au electrodes, showed strong evidence of Fowler–Nordheim tunnelling [29.230]. Since these films were relatively thick, high voltages could be applied, which reduced the Fermi level barrier thickness to that suitable for

tunnelling. A linear dependence of $\log J/V^2$ on $1/V$ was obtained, but the measured currents were somewhat lower than those predicted by the theory, probably as the result of trapping effects. Effective masses of $0.48m$ for Ag, $0.39m$ for Al and $0.42m$ for Si were estimated, where m is the free electron mass. Since these early tunnelling measurements, the mechanism has been observed in many different insulating films.

Tunnelling in semiconductors forms the basis of the operation of the tunnel diode. However, in this case a doubly degenerate p–n junction is used, and tunnelling occurs across the depletion region. The operation of this device is described in a standard text on semiconductor devices [29.231]. In sandwich samples of nondegenerate materials, tunnelling has been observed in CdTe [29.202]. In this case there was a linear dependence of the logarithm of the drift mobility on reciprocal electric field for fields in excess of 1.2×10^6 V/m, which is consistent with (29.109). Tunnelling has also been observed in sandwich structures consisting of a copper phthalocyanine film with In electrodes [29.232]. A modified Fowler–Nordheim expression for tunnelling through an interfacial barrier, when the electric field is sufficient to reduce the barrier width to about 5 nm, has been proposed [29.223], and was applied in this case to the In/Cu phthalocyanine interfacial region. Depending on the sample structure, barrier heights of 0.27 eV and 0.36 eV were derived. Current densities were somewhat lower than predicted by theory, but were accounted for by the fact that tunnelling areas are often considerably lower than the geometric area [29.18] and that in the presence of traps and space charge the current density is also reduced [29.233].

The second electrode-limited conduction process to be considered may occur when the insulating or semiconducting film is too thick for tunnelling to take place and the concentration of allowed states in the forbidden gap is too low for the hopping process to occur. It is known as the Schottky effect, and is the field-assisted lowering of a potential barrier at the injecting electrode. It is similar in origin to the Poole–Frenkel effect, the latter effect having previously been termed the bulk analogue of the Schottky effect [29.234]. The potential barrier at the injecting electrode interface is reduced by an amount $\Delta\phi_S$ which is given by

$$\Delta\phi_S = \beta_S F^{1/2}, \quad (29.110)$$

where

$$\beta_S = \left(\frac{e^3}{4\pi\epsilon_r\epsilon_0} \right)^{1/2} \quad (29.111)$$

is the Schottky field-lowering coefficient. The barrier lowering process is similar to that for the Poole–Frenkel effect, and is also illustrated in various literature reviews [29.164–166]. Differences between the standard Poole–Frenkel and Schottky coefficients given by (29.104) and (29.111) are related to the dissimilar symmetry of the potential barriers in the two cases. Clearly the two field-lowering coefficients are related, with $\beta_{PF} = 2\beta_S$. The fundamental thermionic emission equation of Richardson

$$J = AT^2 \exp\left(-\frac{\phi}{kT}\right) \quad (29.112)$$

gives the current density flowing by a process of electron emission over a potential barrier of height ϕ at a temperature T . In this expression, the constant $A = 1.2 \times 10^6$ A/m², and is known as the Richardson constant. The current density depends only on the potential barrier height and the temperature, and does not require an applied voltage to flow, the electrons acquiring thermal energy only. For the case of Schottky emission, if ϕ_0 , given by (29.91), is the zero-voltage barrier height then the reduced barrier height ϕ is given by $(\phi_0 - \Delta\phi_S)$, so that (29.112) becomes

$$J = AT^2 \exp\left(-\frac{\phi_0}{kT}\right) \exp\left(\frac{\beta_S F^{1/2}}{kT}\right) \quad (29.113)$$

or

$$J = AT^2 \exp\left(-\frac{\phi_0}{kT}\right) \exp\left(\frac{\beta_S V^{1/2}}{kTd^{1/2}}\right). \quad (29.114)$$

Thus, the Schottky effect should show a linear dependence of $\log J$ on $V^{1/2}$, as for the Poole–Frenkel effect described by (29.106). In principle the two effects can be distinguished by the measured value of the field-lowering coefficient, which should be twice as high for the Poole–Frenkel effect as for the Schottky effect.

The Schottky effect has been reported in many inorganic materials, both insulators and semiconductors. *Emtage* and *Tantraporn* [29.235] observed the effect in both Al₂O₃ and GeO₂ sandwich structures having Al electrodes, while *Pollack* [29.236] made similar observations in Al₂O₃ films with Pb electrodes. In this case, not only the constant voltage J – V characteristics were as predicted by (29.114), but the gradients of $\log J$ – $V^{1/2}$ curves increased linearly with reciprocal temperature for temperatures above 235 K, as also predicted by this expression. Below this temperature the variation with temperature was very small, and therefore tunnelling was identified in this region. The dependence of the Schottky effect on the species of electrode was determined in Ta–Ta₂O₅–Au samples prepared by electron

beam evaporation and plasma oxidation [29.237]. In this system the Schottky effect was observed when the Ta electrode was biased negatively and tunnelling when it was biased positively. Similar results were obtained in CdS films [29.169] using In (ohmic) and Au (blocking) contacts. For electron injection from the Au electrode Schottky-type behaviour was observed, with a Schottky barrier height of 0.88 eV obtained from differential capacitance measurements, close to the value expected for Au contacts to CdS single crystals. Field-lowering behaviour has also been observed in Cd₃As₂ films with a bottom Ag electrode and a top electrode of Ag, Al or Au [29.204]. Although the Schottky barrier height is unlikely to be consistent with the simple model of (29.91) owing to the very narrow band gap and the presence of surface states, the experimental value of β was consistent with the value of β_S predicted by (29.111) for a relative permittivity value of 12.

The Schottky effect has also been observed in several organic thin-film materials. Examples are given here only for the phthalocyanines, which are invariably p-type semiconductors. The Schottky effect occurs for hole injection at the blocking contact. Both metal-free [29.238] and copper phthalocyanine films [29.239] having Al electrodes have shown this effect. Behaviour consistent with the Schottky effect has also been observed in planar copper phthalocyanine films with Al electrodes [29.240], with a depletion region width of thickness 96 nm. In sandwich structures having a copper phthalocyanine film and a positively biased Al electrode, values of β several times higher than the theoretical

cal $\beta_S = 2 \times 10^{-5} \text{ eVm}^{1/2}\text{V}^{-1/2}$ were obtained [29.205]. However, if the theoretical value of β_S were assumed and also that most of the voltage was dropped over a depletion region of thickness d_S as suggested previously [29.240], values of $d_S = 120 \text{ nm}$ and $\varphi_0 = 0.88 \text{ eV}$ were derived. This type of assumption has also been subsequently applied in the case of hole injection from a Pb electrode into copper phthalocyanine films [29.175] and for hole injection in triclinic lead phthalocyanine films from an Al electrode [29.177, 206]. In both cases the Schottky barrier width was 50 nm with a barrier height of either 1.11 eV or 1.0 eV.

In this section the main high-field conduction processes that might be observed in insulating and semi-conducting films have been reviewed, and illustrative examples quoted from the literature. It should be emphasised that identification of a particular mechanism is nontrivial, particularly in distinguishing between the Poole–Frenkel and Schottky effects which have very similar J – V dependencies. Furthermore J – V data obtained over a specified voltage range may also show a good correlation with both SCLC and the Poole–Frenkel effect. It is essential to measure temperature dependencies and also to determine the effects of electrode species on the conductivity, making a full consideration of the type of contacts (ohmic or blocking). As devices and structures become ever smaller with the advance of nanotechnology, some of these conduction processes will inevitably become important in nanoelectronic devices. A discussion concerning this aspect is given elsewhere in the literature [29.11].

29.4 Concluding Remarks

In this chapter, an overview of the deposition methods, structures and major properties of thin films has been given. In this section these are briefly summarised and some possible future developments outlined.

Films may be prepared by many different methods. However, these may usually be classified into chemical or physical methods. In the physical methods the film is deposited from a vapour, maintained at a pressure considerably below that of the atmosphere. Vacuum evaporation is the first of these, where the material is transformed into the vapour phase by heating at reduced pressure. Various types of process have been developed to ensure stoichiometry of the deposited material, such as reactive evaporation and co-evaporation. Taken to its limit of sophistication, the method may be operated un-

der UHV conditions using several sources, each of which may be independently controlled. In this case epitaxial films may be deposited, and the method is known as molecular-beam epitaxy (MBE). The second widely used physical method is that of sputtering, which involves the removal of particles (atoms, molecules or larger fragments) from the surface of a target. This is normally achieved using a gas discharge, which delivers high-energy ions to the target surface. Sputtering has several advantages over evaporation, including deposition of stoichiometric materials, better adhesion to the substrate and the capability of depositing most materials. There are several varieties of sputtering, including DC (diode) sputtering, reactive sputtering and RF sputtering, the latter enabling the deposition of insulators as

well as conducting materials. Reactive and co-sputtering give greater flexibility, and higher deposition rates are possible using magnetron sputtering. Various chemical methods may also be utilised for the deposition of films from materials having specific chemical properties. Electrodeposition and electroplating involve passing a high electric current through a solution containing the material, while chemical vapour deposition (CVD) entails the use of various chemical reactions to generate a vapour. The latter method is frequently used in the deposition of electronic materials, and substantial work has been performed in developing processes for several of these. Again there are several different subdivisions of this method, operating at different pressures and with different types of precursor materials. A third chemical method that has found favour in recent years is Langmuir–Blodgett deposition, where molecular layers are applied individually using a dipping process. It has the advantage that high vacuum and elevated temperatures are not required, and film thicknesses are easily derived from the number of layers deposited. It is, however, limited to certain defined materials, and does not have the flexibility of the other methods.

In common with bulk materials, the crystal structure and morphology of thin films may be investigated using diffraction methods and various microscopic techniques (electron microscopy, scanning tunnelling microscopy, etc.). Diffraction methods allow the type of crystal structure to be determined by measurement of the unit cell dimensions in the case of epitaxial and polycrystalline films. Frequently a material may exist in two or more crystalline modifications, and generally these may be distinguished by comparison of diffraction data from the film with tabulated standard data. The type of crystalline modification that is deposited can depend sensitively on the deposition conditions such as the substrate temperature and deposition rate, and this dependence may be established by systematically depositing films under a wide combination of conditions. Many films have a preferred crystalline orientation, and this too may sometimes be correlated with deposition conditions and the substrate morphology. In polycrystalline films, the crystallites are generally preferentially oriented and show a fibre texture. Estimates of the mean microcrystallite size may be obtained from the width of the diffraction peaks using the Scherrer method. The morphological characteristics of thin films commonly have a columnar grain structure when physical deposition methods are used, and the orientation of the grains may sometimes be related to the incident direction of the depositing atoms or molecules. Larger-scale features of

the morphology may often be accounted for in terms of a zone model, where the observed features are identified with the ratio of the substrate temperature to the melting temperature of the film material. Examples of some of these effects in different thin-film materials have been given, and although these are representative of behaviour shown in many thin-film systems, it is clear that both crystal structure and morphology require determination in any specific process used to deposit a particular film material. These features are likely to be particularly important in influencing the optical and electrical properties.

The optical and electrical properties of thin films are those most relevant to this Handbook, and examples of some of the basic properties have been considered as a foundation for some of the more advanced applications described in other chapters. The interference properties of light are fundamental to most of the phenomena observed. In principle the properties of light propagating through media are described by Maxwell's equations, which allow a full analysis for both nonabsorbing and absorbing films for various states of polarisation and optical frequencies. The Fresnel coefficients for reflection and transmission of light at the boundaries between different optical media allow expressions for the energy reflectance and transmittance to be derived. Some examples of these have been given, together with an introduction to matrix methods which may be used to calculate the electric and magnetic field vectors for optical designs of considerable complexity, involving both nonabsorbing and absorbing films. The operation of optical filters, other optical components and fibre-optic waveguides may be predicted by advanced application of these methods. Examples of the basic electrical properties of thin films, which are directly related to the film thickness, are lateral electrical conductivity and high-field conduction processes. Lateral electrical conduction in thin films is a very important feature, which is determined primarily by electron scattering within the film. For metallic films, conductivity–film thickness relationships have been derived in terms of a parameter k , the ratio of the film thickness to the bulk electron mean free path. One would intuitively expect the conductivity to decrease (or the resistivity to increase) with decreasing k , since this implies a curtailment of the mean free path. This problem was first tackled over a century ago by Thomson, and more realistic expressions have been obtained subsequently by solving the Boltzmann transport equation. Unfortunately, the derived general expressions are somewhat complex, but a wide variety of approximations have been obtained for various ranges of k ,

and the conductivity is predicted to reduce to that of the bulk material for thick films. Further refinements in the analysis were made by introducing a specularly parameter, p , which takes into account the proportion of electrons which are specularly scattered at the surfaces. Better agreement with experimental results is generally obtained if the value of p is intermediate between that for totally diffuse scattering ($p = 0$) and totally specular scattering ($p = 1$). Another development of the theory includes assigning different specularly parameters to the two film surfaces, since scattering at the substrate surface is clearly different from scattering at the top surface. In general the various expressions derived from this type of model are consistent with each other, in that those derived from the more complex assumptions normally reduce to those obtained using simpler assumptions if suitable limits are chosen. Scattering by internal grain boundaries has also been investigated by solving the Boltzmann transport equation, and allows contributions to the conductivity resulting from internal features, as well as the surfaces, to be included in the analysis. The ratio of the thin-film to bulk temperature coefficient of resistivity has been shown to be equal to the ratio of thin-film to bulk conductivity for the case of thicker films, although no such convenient relationship applies for thinner films owing to the dependence of the bulk mean free path on temperature, which is particularly significant for very thin films. Considerations of these types of effect need to be fully assessed when designing interconnects to integrated circuits operating in particular temperature ranges. In insulating or semiconducting thin films, very high electric fields may be applied across the thickness of the film. Under these circumstances a variety of different conduction processes may occur, none of which are observed at lower fields. Crucial to the type of conduction process which develops is the type of contact at the interface. Ohmic contacts allow currents to pass unimpeded by the contact, and the conduction process is bulk-limited, while blocking contacts effectively limit the conductivity, which is then electrode-limited. Bulk-limited conduction processes include SCLC, Poole–Frenkel conductivity and hopping. Various types of SCLC have been observed, depending on the type of trap distribution present, if any, and may be distinguished by their characteristic voltage and temperature dependencies. Poole–Frenkel conductivity results from field-dependent lowering of the potential barrier for emission of electrons into the conduction band. Again, various varieties of this process have been observed depending on the appearance of donors and/or traps, and also for the case of nonuniform

fields. Hopping is observed particularly in noncrystalline solids, where electrons hop from one localised energy level to another. It is characterised by low mobility, but requires little thermal energy, and is therefore often observed at low temperatures. Different types of hopping are characterised by the mean distance of the electron hopping process, and show different characteristic behaviours. Electrode-limited conduction processes include tunnelling and the Schottky effect. Tunnelling is a quantum-mechanical effect, where for a thin barrier the electron wavefunction has a finite probability of existence on the opposite side of the barrier. Tunnelling may be either directly from one electrode to the other, or may be from the Fermi level of the injecting electrode into the conduction band of the insulator or semiconductor. The type of tunnelling depends on the relative values of the applied voltage and the barrier height at the interface. The tunnelling current is almost independent of temperature, a feature which allows it to be identified relatively easily. For structures where the barrier is too thick for tunnelling to occur, the Schottky effect may take place instead. This is the field-assisted lowering of a potential barrier at the interface, and has several similarities with the bulk Poole–Frenkel effect. Similar J – V characteristics are predicted for both, but may usually be distinguished from the value of a measured field-lowering coefficient, or from considerations of the current level drawn with varying electrode species.

In conclusion, it should be noted that in this chapter only the main established deposition methods and some of the structural, optical and electrical properties of thin films have been reviewed. Although it is hoped that the references included herein are sufficient to allow the reader to explore the subject in some considerable depth, it is clear that these can only provide a snapshot of knowledge at the present time. During the past half century, there have been tremendous developments in all areas of thin-film knowledge, although the basic principles and properties stressed in this chapter underpin most of them. The main physical methods of deposition have been known for a century or so, but improvements have been contingent on advances in vacuum technology, gas-handling capabilities and theoretical work concerning atomic interactions in solids. MBE deposition systems of considerable complexity and sophistication now exist, enabling the deposition of a considerable range of materials in a variety of structures and designs, although only some of the most important processes have been developed to the level where they are employed in automated manufacture. Control of many factors are vital in successful deposi-

tion processes, and progress in this area has gone hand in hand with the development of computerised monitoring and data-handling systems. Progress in the manipulation of individual atoms and molecules has been made by the adoption of techniques originally developed for very high-resolution surface imaging, such as scanning tunnelling and atomic force microscopy. Chemical methods of film deposition have developed into many dedicated CVD systems, exploiting reactions which are specific to a particular material or class of materials. Additionally many types of films may be deposited using the Langmuir–Blodgett technique, although this relies on the ingenuity of chemists to develop suitable substituted organic compounds. Further development of many of these deposition methods will undoubtedly progress in future years, particularly as the power of computer systems increases. The most significant advances are likely to be in the area of nanotechnology, where MBE systems, equipped with facilities to build nanostructures atom by atom, are currently under development. The structure and morphology adopted by thin films are likely to have a significant influence on all their properties; this is not restricted to just the optical and electrical properties described in this chapter. Diffraction methods are becoming more precise, and can now be used for the investigation of real-time changes when synchrotron radiation is used. Scanning tunnelling and atomic force

microscopic techniques have enabled the acquisition of surface images with resolution of atomic dimensions, and the increased use and development of these and associated techniques will be necessary for future investigation of nanostructures. Optical properties described in this chapter will no doubt be exploited in more complex and sophisticated filter and mirror designs, as well as their employment in fibre-optic cables and applications in photonics. Recent developments in the area of negative-refractive-index materials will no doubt also be exploited in due course. The electrical properties of thin films under both low- and high-field conditions are reasonably well understood, and it is anticipated that applications are likely to be made in the area of nanoelectronics. Some possible applications have been proposed elsewhere [29.11]. The provision of electrical contacts will remain of particular importance, although it may not be possible to consider the contact region separately from the remaining nanostructure. The efficient injection of charge into nanostructures using tunnelling and Schottky barrier (blocking) contacts may be achieved, and high current densities may be obtained with trap-free SCLC contacts. It is nevertheless clear, however, that in all areas of future development in thin-film technology, the control of deposition processes and materials processing aspects are of paramount importance. This is likely to remain a considerable challenge in the future.

References

- 29.1 L. Eckertová: *Physics of Thin Films*, 2nd edn. (Plenum, New York 1986)
- 29.2 J. Thewlis: *Concise Dictionary of Physics and Related Subjects*, 2nd edn. (Pergamon, Oxford 1979) p. 336
- 29.3 R. W. Hoffman: The mechanical properties of thin condensed films. In: *Physics of Thin Films*, Vol. 3, ed. by G. Hass, R. E. Thun (Academic, San Diego 1966) p. 211
- 29.4 D. S. Campbell: Mechanical properties of thin films. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw–Hill, New York 1970) Chap. 12
- 29.5 C. A. Neugebauer, M. B. Webb: *J. Appl. Phys.* **33**, 74 (1962)
- 29.6 L. I. Maissel: Electronic properties of metallic thin films. In: *Handbook of Thin Film Technology*, Vol. 13, ed. by L. I. Maissel, R. Glang (McGraw–Hill, New York 1970)
- 29.7 K. H. Behrndt: Film-thickness and deposition-rate monitoring devices and techniques for producing films of uniform thickness. In: *Physics of Thin Films*, Vol. 3, ed. by G. Hass, R. E. Thun (Academic, San Diego 1966.) p. 1
- 29.8 W. A. Pliskin, S. J. Zanin: Film thickness and composition. In: *Handbook of Thin Film Technology*, Vol. 11, ed. by L. I. Maissel, R. Glang (McGraw–Hill, New York 1970)
- 29.9 J. P. Hirth, K. L. Moazed: Nucleation processes in thin film formation. In: *Physics of Thin Films*, Vol. 4, ed. by G. Hass, R. E. Thun (Academic, San Diego 1967) p. 97
- 29.10 C. A. Neugebauer: *Condensation, Nucleation and Growth of Thin Films*, Vol. 8, ed. by L. I. Maissel, R. Glang (McGraw–Hill, New York 1970)
- 29.11 R. D. Gould: High field conduction in nanostructures. In: *Encyclopedia of Nanoscience and Nanotechnology*, Vol. 3, ed. by H. S. Nalwa (American Scientific, Stevenson Ranch 2004) pp. 891–915
- 29.12 *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw–Hill, New York 1970)
- 29.13 *Thin Films*, Vol. 25, ed. by S. M. Rossmagel, A. Ullman, M. H. Francombe (Academic, San Diego 1998.)

- 29.14 *Handbook of Thin Film Materials*, ed. by H. S. Nalwa (Academic, San Diego 2001)
- 29.15 O. S. Heavens: *Thin Film Physics* (Methuen, London 1970)
- 29.16 L. I. Maissel, M. H. Francombe: *An Introduction to Thin Films* (Gordon and Breach, New York 1973)
- 29.17 *Active and Passive Thin Film Devices*, ed. by T. J. Coutts (Academic, New York 1978)
- 29.18 K. I. Chopra: *Thin Film Phenomena* (Krieger, New York 1979) (first published by McGraw-Hill, New York, 1969)
- 29.19 M. Ohring: *The Materials Science of Thin Films* (Academic, San Diego 1992)
- 29.20 R. E. Honig, D. A. Kramer: *RCA Rev.* **30**, 285 (1969)
- 29.21 *Scientific Foundation of Vacuum Technique*, 2nd edn., ed. by S. Dushman, J. M. Lafferty (Wiley, New York 1962)
- 29.22 R. Glang: Vacuum evaporation. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 1
- 29.23 O. Kubaschewski, S. L. Evans, C. B. Alcock: *Metallurgical Thermochemistry*, 4th revised edn. (Pergamon, Oxford 1967)
- 29.24 R. E. Honig: *RCA Rev.* **23**, 567 (1962)
- 29.25 H. Hertz: *Ann. Phys. Chem.* **17**, 177 (1882)
- 29.26 M. Knudsen: *Ann. Phys.* **47**, 697 (1915)
- 29.27 J. P. Hirth, G. M. Pound: *Condensation and Evaporation* (Pergamon, Oxford 1963)
- 29.28 L. Holland, W. Steckelmacher: *Vacuum* **2**, 346 (1952)
- 29.29 A. von Hippel: *Ann. Phys.* **81**, 1043 (1926)
- 29.30 H. A. Macleod: *Thin-Film Optical Filters* (Adam Hilger, London 1969)
- 29.31 T. C. Tisone, J. B. Bindell: *J. Vac. Sci. Technol.* **11**, 72 (1974)
- 29.32 R. Brown: Thin film substrates. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 6
- 29.33 G. Sauerbrey: *Z. Phys.* **155**, 206 (1959)
- 29.34 C. D. Stockbridge: Resonance frequency versus mass added to quartz crystals. In: *Vacuum Microbalance Techniques*, Vol. 5, ed. by K. H. Behrndt (Plenum, New York 1966) p. 193
- 29.35 C. J. Bowler, R. D. Gould: *J. Vac. Sci. Technol. A* **5**, 114 (1987)
- 29.36 R. Hultgren, R. L. Orr, P. D. Anderson, K. K. Kelley: *Selected Values of Thermodynamic Properties of Metals and Alloys* (Wiley, New York 1963)
- 29.37 G. Zinsmeister: *Vakuum-Tech.* **8**, 223 (1964)
- 29.38 R. D. Gould: Electrical conduction properties of thin films of cadmium compounds. In: *Handbook of Thin Film Materials*, ed. by H. S. Nalwa (Academic, San Diego 2001) Chap. 4, pp. 187–245
- 29.39 D. R. Stull: *JANAF Thermochemical Tables* (Dow Chemical Co., U.S. Clearinghouse, Springfield, Virginia 1970)
- 29.40 F. A. Pizzarello: *J. Appl. Phys.* **35**, 2730 (1964)
- 29.41 P. A. Timson, C. A. Hogarth: *Thin Solid Films* **8**, 237 (1971)
- 29.42 A. Y. Cho: *Thin Solid Films* **100**, 291 (1983)
- 29.43 B. A. Joyce: *Rep. Prog. Phys.* **48**, 1637 (1985)
- 29.44 Y. Ota: *Thin Solid Films* **106**, 3 (1983)
- 29.45 G. K. Wehner, G. S. Anderson: The nature of physical sputtering. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 3
- 29.46 O. Almén, G. Bruce: *Nucl. Instrum. Methods* **11**, 279 (1961)
- 29.47 F. Keywell: *Phys. Rev.* **97**, 1611 (1955)
- 29.48 P. Sigmund: Sputtering by ion bombardment: Theoretical concepts. In: *Sputtering by Particle Bombardment I*, ed. by R. Behrisch (Springer, Berlin, Heidelberg 1981) Chap. 2, p. 9
- 29.49 R. S. Pease: *Rendiconti della Scuola Internazionale di Fisica "Enrico Fermi"*, Corso XIII (Società Italiana di Fisica, Bologna 1959) p. 158
- 29.50 P. Sigmund: *Phys. Rev.* **184**, 383 (1969)
- 29.51 L. I. Maissel: The deposition of thin films by cathode sputtering. In: *Physics of Thin Films*, Vol. 3, ed. by G. Hass, R. E. Thun (Academic, San Diego 1966) p. 61
- 29.52 G. H. Kinchin, R. S. Pease: *Rep. Prog. Phys.* **18**, 1 (1955)
- 29.53 H. H. Andersen, H. L. Bay: Sputtering yield measurements. In: *Sputtering by Particle Bombardment I*, ed. by R. Behrisch (Springer, Berlin, Heidelberg 1981) Chap. 4, p. 145
- 29.54 J. Lindhard, V. Nielsen, M. Scharff: *Kgl. Danske Vidensk. Selskab, Mat.-Fys. Medd.* **36**(10), 1 (1968)
- 29.55 A. von Engel: *Ionized Gases*, 2nd edn. (Clarendon, Oxford 1965)
- 29.56 F. Llewellyn-Jones: *The Glow Discharge* (Methuen, London 1966)
- 29.57 L. I. Maissel: Applications of sputtering to the deposition of films. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 4
- 29.58 G. Perny, M. Samirant, B. Laville Saint Martin: *Compt. Rend. Acad. Sci. Ser. C* **262**, 265 (1966)
- 29.59 G. Perny, B. Laville Saint Martin: Proceedings of the International Symposium on Basic Problems in Thin Film Physics, Clausthal 1965, ed. by R. Niedermayer, H. Mayer (Vandenhoeck and Ruprecht, Göttingen 1966) 709
- 29.60 S. A. Awan, R. D. Gould: *Thin Solid Films* **423**, 267 (2003)
- 29.61 W. D. Westwood: Reactive sputtering. In: *Physics of Thin Films*, Vol. 14, ed. by M. H. Francombe, J. C. Vossen (Academic, San Diego 1989) p. 1
- 29.62 W. R. Sinclair, F. G. Peters: *Rev. Sci. Instrum.* **33**, 744 (1992)
- 29.63 L. I. Maissel, P. M. Schaible: *J. Appl. Phys.* **36**, 237 (1965)
- 29.64 R. Frerichs: *J. Appl. Phys.* **33**, 1898 (1962)
- 29.65 H. C. Theuerer, J. J. Hauser: *J. Appl. Phys.* **35**, 554 (1964)
- 29.66 E. W. Williams: *The CD-ROM and Optical Disc Recording Systems* (Oxford Univ. Press, Oxford 1996)

- 29.67 H. E. Winters, E. Kay: J. Appl. Phys. **38**, 3928 (1967)
- 29.68 G. S. Anderson, W. N. Mayer, G. K. Wehner: J. Appl. Phys. **33**, 2991 (1962)
- 29.69 P. D. Davidse, L. I. Maissel: J. Appl. Phys. **37**, 574 (1966)
- 29.70 R. A. Powell, S. M. Rossnagel: *PVD for Microelectronics: Sputter Deposition Applied to Semiconductor Manufacturing*, Thin Films, Vol. 26 (Academic, San Diego 1999) p. 51
- 29.71 J. S. Logan, N. M. Mazza, P. D. Davidse: J. Vac. Sci. Technol. **6**, 120 (1969)
- 29.72 S. Rohde: Unbalanced magnetron sputtering. In: *Physics of Thin Films*, Vol. 18, ed. by M. H. Francombe, J. L. Vossen (Academic, San Diego 1994) p. 235
- 29.73 B. Window, N. Savvides: J. Vac. Sci. Technol. A **4**, 196 (1986)
- 29.74 Y. Ochiai, K. Aso, M. Hayakawa, H. Matsuda, K. Hayashi, W. Ishikawa, Y. Iwasaki: J. Vac. Sci. Technol. A **4**, 19 (1986)
- 29.75 D. S. Campbell: The deposition of thin films by chemical methods. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 5
- 29.76 D. S. Campbell: Preparation methods for thin film devices. In: *Active and Passive Thin Film Devices*, ed. by T. J. Coutts (Academic, New York 1978) p. 23
- 29.77 A. Brenner: *Electrodeposition of Alloys*, Vol. 1, 2 (Academic, New York 1963)
- 29.78 K. R. Lawless: The growth and structure of electrodeposits. In: *Physics of Thin Films*, Vol. 4, ed. by G. Hass, R. E. Thun (Academic, San Diego 1967) p. 191
- 29.79 L. Young: *Anodic Oxide Films* (Academic, New York 1961)
- 29.80 C. J. Dell'Oca, D. L. Pulfrey, L. Young: Anodic oxide films. In: *Physics of Thin Films*, Vol. 6, ed. by M. H. Francombe, R. W. Hoffman (Academic, San Diego 1971) p. 1
- 29.81 W. M. Feist, S. R. Steele, D. W. Readey: The preparation of films by chemical vapor deposition. In: *Physics of Thin Films*, Vol. 5, ed. by G. Hass, R. E. Thun (Academic, San Diego 1969) p. 237
- 29.82 J. R. Knight, D. Effer, P. R. Evans: Solid State Electron. **8**, 178 (1965)
- 29.83 M. H. Francombe, J. E. Johnson: The preparation and properties of semiconductor films. In: *Physics of Thin Films*, Vol. 5, ed. by G. Hass, R. E. Thun (Academic, San Diego 1969) p. 143
- 29.84 S. M. Ojha: Plasma-enhanced chemical vapor deposition of thin films. In: *Physics of Thin Films*, Vol. 12, ed. by G. Hass, M. H. Francombe, J. L. Vossen (Academic, San Diego 1982) p. 237
- 29.85 A. C. Adams: Dielectric and polysilicon film deposition. In: *VLSI Technology*, 2nd edn., ed. by S. M. Sze (McGraw-Hill, New York 1988) Chap. 6, p. 233
- 29.86 O. A. Popov: Electron cyclotron resonance plasma sources and their use in plasma-assisted chemical vapor deposition of thin films. In: *Physics of Thin Films*, Vol. 18, ed. by M. H. Francombe, J. L. Vossen (Academic, San Diego 1994) p. 121
- 29.87 C. W. Pitt, L. M. Walpitta: Thin Solid Films **68**, 101 (1980)
- 29.88 N. G. Dhere: High- T_c superconducting thin films. In: *Physics of Thin Films*, Vol. 16, ed. by M. H. Francombe, J. L. Vossen (Academic, San Diego 1992) p. 1
- 29.89 I. Langmuir: Trans. Faraday Soc. **15**, 62 (1920)
- 29.90 K. B. Blodgett: J. Am. Chem. Soc. **56**, 495 (1934)
- 29.91 G. L. Gaines: Thin Solid Films **99**, ix (1983)
- 29.92 S. Baker, M. C. Petty, G. G. Roberts, M. V. Twigg: Thin Solid Films **99**, 53 (1983)
- 29.93 M. C. Petty: *Langmuir-Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge 1996)
- 29.94 L. S. Miller, P. J. W. Stone: Thin Solid Films **210/211**, 19 (1992)
- 29.95 L. S. Miller, A. L. Rhoden: Thin Solid Films **243**, 339 (1994)
- 29.96 P. S. Vincett, G. G. Roberts: Thin Solid Films **68**, 135 (1980)
- 29.97 A. N. Broers, M. Pomerantz: Thin Solid Films **99**, 323 (1983)
- 29.98 R. E. Thun: Structure of thin films. In: *Physics of Thin Films*, Vol. 1, ed. by G. Hass (Academic, San Diego 1963) p. 187
- 29.99 C. A. Neugebauer: Structural disorder phenomena in thin metal films. In: *Physics of Thin Films*, Vol. 2, ed. by G. Hass, R. E. Thun (Academic, San Diego 1964) p. 1
- 29.100 I. H. Khan: The growth and structure of single-crystal films. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 10
- 29.101 G. B. Stringfellow: Rep. Prog. Phys. **43**, 469 (1982)
- 29.102 D. Walton: J. Chem. Phys. **37**, 2182 (1962)
- 29.103 D. Walton, T. N. Rhodin, R. Rollins: J. Chem. Phys. **38**, 2695 (1963)
- 29.104 K. K. Muravjeva, I. P. Kalinkin, V. B. Aleskowsky, N. S. Bogomolov: Thin Solid Films **5**, 7 (1970)
- 29.105 V. N. E. Robinson, J. L. Robins: Thin Solid Films **20**, 155 (1974)
- 29.106 J. A. Venables, G. D. T. Spiller: Rep. Prog. Phys. **47**, 399 (1984)
- 29.107 I. P. Kalinkin, K. K. Muravyeva, L. A. Sergeyewa, V. B. Aleskowsky, N. S. Bogomolov: Krist. Tech. **5**, 51 (1970)
- 29.108 D. B. Holt: Thin Solid Films **24**, 1 (1974)
- 29.109 K. Zanio: Cadmium telluride. In: *Semiconductors and Semimetals*, Vol. 13, ed. by R. K. Willardson, A. C. Beer (Academic, New York 1978)
- 29.110 A. Ashour, R. D. Gould, A. A. Ramadan: Phys. Status Solidi A **125**, 541 (1991)
- 29.111 R. D. Gould: Coord. Chem. Rev. **156**, 237 (1996)
- 29.112 A. K. Hassan, R. D. Gould: Phys. Status Solidi A **132**, 91 (1992)
- 29.113 S. I. Shihub, R. D. Gould: Phys. Status Solidi A **139**, 129 (1993)
- 29.114 P. Scherrer: Gott. Nachr. **2**, 98 (1918)

- 29.115 M. Ashida, N. Uyeda, E. Suito: Bull. Chem. Soc. Jpn. **39**, 2616 (1966)
- 29.116 L. Źdanowicz, S. Miotkowska: Thin Solid Films **29**, 177 (1975)
- 29.117 G. A. Steigmann, J. Goodyear: Acta Crystallogr. B **24**, 1062 (1968)
- 29.118 A. Bienenstock, F. Betts, S. R. Ovshinsky: J. Non-Cryst. Solids **2**, 347 (1970)
- 29.119 S. R. Elliott: *Physics of Amorphous Materials*, 2nd edn. (Longman, London 1990)
- 29.120 J. R. Bosnell: Amorphous semiconducting films. In: *Active and Passive Thin Film Devices*, ed. by T. J. Coutts (Academic, New York 1978) p. 245
- 29.121 N. F. Mott, E. A. Davis: *Electronic Processes in Non-Crystalline Materials*, 2nd edn. (Oxford Univ. Press, Oxford 1979)
- 29.122 D. B. Dove: Electron diffraction analysis of the local atomic order in amorphous films. In: *Physics of Thin Films*, Vol. 7, ed. by G. Hass, M. H. Francombe, R. W. Hoffman (Academic, San Diego 1973) p. 1
- 29.123 K. H. Norian, J. W. Edington: Thin Solid Films **75**, 53 (1981)
- 29.124 J. M. Nieuwenhuizen, H. B. Haanstra: Philips Tech. Rev. **27**, 87 (1966)
- 29.125 S. B. Hussain: Thin Solid Films **22**, S5 (1974)
- 29.126 J. I. B. Wilson, J. Woods: J. Phys. Chem. Solids **34**, 171 (1973)
- 29.127 A. G. Stanley: Cadmium sulphide solar cells. In: *Applied Solid State Science*, Vol. 5, ed. by R. Wolfe (Academic, New York 1975) p. 251
- 29.128 B. Goldstein, L. Pensak: J. Appl. Phys. **30**, 155 (1959)
- 29.129 J. Saraie, M. Akiyama, T. Tanaka: Jpn. J. Appl. Phys. **11**, 1758 (1972)
- 29.130 Yu. K. Yezhovskiy, I. P. Kalinkin: Thin Solid Films **18**, 127 (1973)
- 29.131 J. Jurusik, L. Źdanowicz: Thin Solid Films **67**, 285 (1980)
- 29.132 K. Tanaka: Jpn. J. Appl. Phys. **9**, 1070 (1970)
- 29.133 T. M. Ratcheva-Stambolieva, Yu. D. Tchistyakov, G. A. Krasulin, A. V. Vanyukov, D. H. Djoglev: Phys. Status Solidi A **16**, 315 (1973)
- 29.134 H. W. Lehmann, R. Widmer: Thin Solid Films **33**, 301 (1976)
- 29.135 J. Jurusik: Thin Solid Films **214**, 117 (1992)
- 29.136 J. Jurusik: Thin Solid Films **248**, 178 (1994)
- 29.137 Z. Knittl: *Optics of Thin Films* (Wiley, New York 1976)
- 29.138 O. S. Heavens: *Optical Properties of Thin Solid Films* (Dover, New York 1965) (first published by Butterworths, London, 1955)
- 29.139 O. S. Heavens: Measurement of optical constants of thin films. In: *Physics of Thin Films*, Vol. 2, ed. by G. Hass, R. E. Thun (Academic, San Diego 1964) p. 193
- 29.140 O. S. Heavens: Optical properties of thin films. In: *Encyclopaedic Dictionary of Physics*, Supplementary, Vol. 3, ed. by J. Thewlis (Pergamon, Oxford 1969) p. 412
- 29.141 H. A. Macleod: Thin film optical devices. In: *Passive and Active Thin Film Devices*, ed. by T. J. Coutts (Academic, New York 1978) p. 321
- 29.142 A. Thelen: Design of multilayer interference filters. In: *Physics of Thin Films*, Vol. 5, ed. by G. Hass, R. E. Thun (Academic, San Diego 1969) p. 47
- 29.143 F. Abelès: Optical properties of metallic thin films. In: *Physics of Thin Films*, Vol. 6, ed. by M. H. Francombe, R. W. Hoffman (Academic, San Diego 1971) p. 151
- 29.144 P. H. Berning: Theory and calculations of optical thin films. In: *Physics of Thin Films*, Vol. 1, ed. by G. Hass (Academic, San Diego 1963) p. 69
- 29.145 J. T. Cox, G. Hass: Antireflection coatings for optical and infrared optical materials. In: *Physics of Thin Films*, Vol. 2, ed. by G. Hass, R. E. Thun (Academic, San Diego 1964) p. 239
- 29.146 G. Hass, M. H. Francombe, R. W. Hoffman: Metal-dielectric interference filters. In: *Physics of Thin Films*, Vol. 9, ed. by G. Hass, M. H. Francombe, R. W. Hoffman (Academic, San Diego 1977) p. 74
- 29.147 J. M. Eastman: Scattering in all-dielectric multilayer bandpass filters and mirrors for lasers. In: *Physics of Thin Films*, Vol. 10, ed. by G. Hass, M. H. Francombe (Academic, San Diego 1978) p. 167
- 29.148 G. Hass, J. B. Heaney, W. R. Hunter: Reflectance and preparation of front surface mirrors for use at various angles of incidence from the ultraviolet to the far infrared. In: *Physics of Thin Films*, Vol. 12, ed. by G. Hass, M. H. Francombe, J. L. Vossen (Academic, San Diego 1982) p. 2
- 29.149 J. N. Zemel: Transport phenomena in hetero-epitaxial semiconductor films. In: *The Use of Thin Films in Physical Investigations*, ed. by J. C. Anderson (Academic, New York 1966) p. 319
- 29.150 K. Fuchs: Proc. Cambridge Phil. Soc. **34**, 100 (1938)
- 29.151 J. J. Thomson: Proc. Cambridge Phil. Soc. **11**, 120 (1901)
- 29.152 A. C. B. Lovell: Proc. R. Soc. A **157**, 311 (1936)
- 29.153 D. S. Campbell: The electrical properties of single-crystal metal films. In: *The Use of Thin Films in Physical Investigations*, ed. by J. C. Anderson (Academic, New York 1966) p. 299
- 29.154 E. H. Sondheimer: Adv. Phys. **1**, 1 (1952)
- 29.155 M. S. P. Lucas: J. Appl. Phys. **36**, 1632 (1965)
- 29.156 H. J. Juretschke: Surface Sci. **2**, 40 (1964)
- 29.157 H. J. Juretschke: J. Appl. Phys. **37**, 435 (1966)
- 29.158 D. C. Larson: Size-dependent electrical conductivity in thin metal films and wires. In: *Physics of Thin Films*, Vol. 6, ed. by M. H. Francombe, R. W. Hoffman (Academic, San Diego 1971) p. 81
- 29.159 A. F. Mayadas, M. Shatzkes: Phys. Rev. B **1**, 1382 (1970)
- 29.160 C. R. Tellier, A. J. Tosser: *Size Effects in Thin Films* (Elsevier Science, Amsterdam 1982)
- 29.161 E. E. Mola, J. M. Heras: Thin Solid Films **18**, 137 (1973)
- 29.162 T. J. Coutts: Electrical properties and applications of thin metallic and alloy films. In: *Active and Passive*

- Thin Film Devices*, ed. by T. J. Coutts (Academic, New York 1978) p. 57
- 29.163 E. E. Mola, J. M. Heras: *Electrocomp. Sci. Technol.* **1**, 77 (1974)
- 29.164 J. G. Simmons: *J. Phys. D: Appl. Phys.* **4**, 613 (1971)
- 29.165 E. H. Rhoderick: *Metal–Semiconductor Contacts* (Clarendon, Oxford 1978)
- 29.166 D. R. Lamb: *Electrical Conduction Mechanisms in Thin Insulating Films* (Metheun, London 1967)
- 29.167 A. Rose: *Phys. Rev.* **97**, 1538 (1955)
- 29.168 T. Budinas, P. Mackus, A. Smilga, J. Vivvakas: *Phys. Stat. Sol.* **31**, 375 (1969)
- 29.169 R. Zuleeg: *Solid State Electron.* **6**, 645 (1963)
- 29.170 R. Zuleeg, R. S. Muller: *Solid State Electron.* **7**, 575 (1964)
- 29.171 B. M. Basol, O. M. Stafsudd: *Solid State Electron.* **24**, 121 (1981)
- 29.172 V. S. Dharmadhikari: *Int. J. Electron.* **54**, 787 (1983)
- 29.173 R. Glew: *Thin Solid Films* **46**, 59 (1977)
- 29.174 R. K. Pandey, R. B. Gore, A. J. N. Rooz: *J. Phys. D: Appl. Phys.* **20**, 1059 (1987)
- 29.175 R. D. Gould, A. K. Hassan: *Thin Solid Films* **193/194**, 895 (1990)
- 29.176 A. K. Hassan, R. D. Gould: *J. Phys: Condens. Matter* **1**, 6679 (1989)
- 29.177 T. S. Shafai, R. D. Gould: *Int. J. Electron.* **73**, 307 (1992)
- 29.178 A. K. Hassan, R. D. Gould: *Int. J. Electron.* **74**, 59 (1993)
- 29.179 S. Gravano, A. K. Hassan, R. D. Gould: *Int. J. Electron.* **70**, 477 (1991)
- 29.180 B. B. Ismail, R. D. Gould: *Phys. Status Solidi A* **115**, 237 (1989)
- 29.181 R. D. Gould, M. S. Rahman: *J. Phys. D: Appl. Phys.* **14**, 79 (1981)
- 29.182 M. A. Lampert: *Rep. Prog. Phys.* **27**, 329 (1964)
- 29.183 R. D. Gould, B. A. Carter: *J. Phys. D: Appl. Phys.* **16**, L201 (1983)
- 29.184 K. N. Sharma, K. Barua: *J. Phys. D: Appl. Phys.* **12**, 1729 (1979)
- 29.185 A. O. Oduor, R. D. Gould: *Thin Solid Films* **270**, 387 (1995)
- 29.186 A. O. Oduor, R. D. Gould: *Thin Solid Films* **317**, 409 (1998)
- 29.187 A. Sussman: *J. Appl. Phys.* **38**, 2738 (1967)
- 29.188 R. D. Gould: *J. Appl. Phys.* **53**, 3353 (1982)
- 29.189 R. D. Gould: *J. Phys. D: Appl. Phys.* **9**, 1785 (1986)
- 29.190 R. D. Gould, B. B. Ismail: *Int. J. Electron.* **69**, 19 (1990)
- 29.191 S. S. Ou, O. M. Stafsudd, B. M. Basol: *Thin Solid Films* **112**, 301 (1984)
- 29.192 G. M. Delacote, J. P. Fillard, F. J. Marco: *Solid State Commun.* **2**, 373 (1964)
- 29.193 M. A. Lampert, P. Mark: *Current Injection in Solids* (Academic, New York 1970)
- 29.194 S. Gogoi, K. Barua: *Thin Solid Films* **92**, 227 (1982)
- 29.195 R. D. Gould, C. J. Bowler: *Thin Solid Films* **164**, 281 (1988)
- 29.196 A. Servini, A. K. Jonscher: *Thin Solid Films* **3**, 341 (1969)
- 29.197 A. K. Jonscher: *Thin Solid Films* **1**, 213 (1967)
- 29.198 A. K. Jonscher, A. A. Ansari: *Phil. Mag.* **23**, 205 (1971)
- 29.199 S. M. Sze: *J. Appl. Phys.* **38**, 2951 (1967)
- 29.200 H. Murray, A. Tosser: *Thin Solid Films* **36**, 247 (1976)
- 29.201 A. Piel, H. Murray: *Thin Solid Films* **44**, 65 (1977)
- 29.202 C. Canali, F. Nava, G. Ottaviani, K. Zanio: *Solid State Commun.* **13**, 1255 (1973)
- 29.203 R. D. Gould, B. B. Ismail: *Vacuum* **50**, 99 (1998)
- 29.204 M. Din, R. D. Gould: *Thin Solid Films* **340**, 28 (1999)
- 29.205 A. K. Hassan, R. D. Gould: *Int. J. Electron.* **69**, 11 (1990)
- 29.206 A. Ahmad, R. A. Collins: *Phys. Status Solidi A* **126**, 411 (1991)
- 29.207 J. Kašpar, I. Emmer, R. A. Collins: *Int. J. Electron.* **76**, 793 (1994)
- 29.208 N. F. Mott: *Metal–Insulator Transitions* (Taylor and Francis, London 1974)
- 29.209 H. Böttger, V. V. Bryksin: *Hopping Conduction in Solids* (VCH, Weinheim 1985)
- 29.210 A. H. Clark: *Phys. Rev.* **154**, 750 (1967)
- 29.211 P. A. Walley, A. K. Jonscher: *Thin Solid Films* **1**, 367 (1968)
- 29.212 C. J. Adkins, S. M. Freake, E. M. Hamilton: *Phil. Mag.* **22**, 183 (1970)
- 29.213 R. D. Gould, B. B. Ismail: *Phys. Status Solidi A* **134**, K65 (1992)
- 29.214 D. S. H. Chan, A. E. Hill: *Thin Solid Films* **35**, 337 (1976)
- 29.215 J. Mycielski: *Phys. Rev.* **123**, 99 (1961)
- 29.216 J. Le Moigne, R. Even: *J. Chem. Phys.* **83**, 6472 (1985)
- 29.217 A. Ahmad, R. A. Collins: *Thin Solid Films* **217**, 75 (1992)
- 29.218 S. A. James, A. K. Ray, J. Silver: *Phys. Status Solidi A* **129**, 435 (1992)
- 29.219 A. M. Saleh, R. D. Gould, A. K. Hassan: *Phys. Status Solidi A* **139**, 379 (1993)
- 29.220 R. D. Gould, A. K. Hassan: *Thin Solid Films* **223**, 334 (1993)
- 29.221 S. I. Shihub, R. D. Gould: *Thin Solid Films* **254**, 187 (1995)
- 29.222 J. G. Simmons: *J. Appl. Phys.* **34**, 1793 (1963)
- 29.223 J. G. Simmons: *Phys. Rev.* **166**, 912 (1968)
- 29.224 R. H. Fowler, L. W. Nordheim: *Proc. R. Soc. A* **119**, 173 (1928)
- 29.225 J. C. Fisher, I. Giaever: *J. Appl. Phys.* **32**, 172 (1961)
- 29.226 R. Holm: *J. Appl. Phys.* **22**, 569 (1951)
- 29.227 D. Meyerhofer, S. A. Ochs: *J. Appl. Phys.* **34**, 2535 (1963)
- 29.228 S. R. Pollack, C. E. Morris: *J. Appl. Phys.* **35**, 1503 (1964)
- 29.229 J. G. Simmons: *J. Appl. Phys.* **34**, 2581 (1964)
- 29.230 M. Lenzlinger, E. H. Snow: *J. Appl. Phys.* **40**, 278 (1969)
- 29.231 S. M. Sze: *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York 1981)

- 29.232 A. K. Hassan, R. D. Gould: J. Phys. D: Appl. Phys. **22**, 1162 (1989)
- 29.233 D. V. Geppert: J. Appl. Phys. **33**, 2993 (1962)
- 29.234 J. G. Simmons: Electronic conduction through thin insulating films. In: *Handbook of Thin Film Technology*, ed. by L. I. Maissel, R. Glang (McGraw-Hill, New York 1970) Chap. 14
- 29.235 P. R. Emtage, W. Tantraporn: Phys. Rev. Lett. **8**, 267 (1962)
- 29.236 S. R. Pollack: J. Appl. Phys. **34**, 877 (1963)
- 29.237 W. E. Flannery, S. R. Pollack: J. Appl. Phys. **37**, 4417 (1966)
- 29.238 M. Fustöss-Wegner: Thin Solid Films **36**, 89 (1976)
- 29.239 B. Sh. Barkhalov, Yu. A. Vidadi: Thin Solid Films **40**, L5 (1977)
- 29.240 A. Wilson, R. A. Collins: Sens. Actuators **12**, 389 (1987)

Thick Films

30. Thick Films

Thick film technology is an example of one of the earliest forms of microelectronics-enabling technologies and it has its origins in the 1950s. At that time it offered an alternative approach to printed circuit board technology and the ability to produce miniature, integrated, robust circuits. It has largely lived in the shadow of silicon technology since the 1960s. The films are deposited by screen printing (stenciling), a graphic reproduction technique that can be dated back to the great Chinese dynasties of around a thousand years ago. Indeed, there is evidence that even early Palaeolithic cave paintings from circa 15 000 BC may have been created using primitive stenciling techniques. With the advent of surface-mounted electronic devices in the 1980s, thick film technology again became popular because it allowed the fabrication of circuits without through-hole components.

This chapter will review the main stages of the thick film fabrication process and discuss some of the commonly used materials and substrates. It will highlight the way in which the technology can be used to manufacture hybrid microelectronic circuits. The latter stages of the chapter will demonstrate how the technology has evolved over the past twenty years or so to become an important method in the production of solid state sensors.

30.1 Thick Film Processing	718
30.1.1 Screen Printing	718
30.1.2 The Drying and Firing Process	719
30.2 Substrates	720
30.2.1 Alumina	720
30.2.2 Stainless Steel	720
30.2.3 Polymer Substrates	720
30.3 Thick Film Materials	721
30.3.1 Conductors	721
30.3.2 Resistors	722
30.3.3 Dielectrics	723
30.3.4 Polymer Thick Films	723
30.4 Components and Assembly	724
30.4.1 Passive Components	724
30.4.2 Active Components	725
30.4.3 Trimming	725
30.4.4 Wire Bonding	726
30.4.5 Soldering	727
of Surface-Mounted Components	727
30.4.6 Packaging and Testing	727
30.5 Sensors	728
30.5.1 Mechanical	728
30.5.2 Thermal	729
30.5.3 Optical	730
30.5.4 Chemical	730
30.5.5 Magnetic	730
30.5.6 Actuators	731
References	731

The term “thick film” is often misinterpreted, and so it is worth noting from the outset that it has little to do with the actual thickness of the film itself. The preferred definition encompasses the fabrication process, namely screen printing, used to deposit the films. The typical range of thicknesses for thick film layers is 0.1 μm to 100 μm . Screen printing is one of the oldest forms of graphic art reproduction and involves the deposition of an ink (or paste) onto a base material (or substrate) through the use of a finely-woven screen with an etched pattern of the desired geometry. The process is commonly used for the production of graphics and

text onto items such as T-shirts, mugs, pencils, textiles and so on, and is very similar to that used for micro-electronic thick films. The degree of sophistication for the latter is, however, significantly greater, resulting in high-quality, reproducible films for use in a variety of electronic systems.

The technology used to manufacture thick film hybrid microelectronic circuits was introduced in the 1950s. Such circuits typically comprised semiconductor devices, monolithic integrated circuits, discrete passive components and the thick films themselves [30.1, 2]. In the early days of the technology, the thick films were

mainly resistors, conductors or dielectric layers. Evidence of thick film circuits can still be found in many of today's commercial devices such as televisions, calculators and telephones. The use of thick film technology was overshadowed in the 1960s by the impact of silicon technology. It found popularity again in the 1980s as a result of the advent of surface-mounted devices, which can be attached to circuit boards using screen-printed solder

layers. Nowadays, thick films are not only used in hybrid circuits but also in advanced solid state sensors and actuators, as we shall see later. In broad terms, thick films can be classified as either *cermets* (ceramic/metallics), requiring high-temperature processing, or *polymer* thick films, which are cured at significantly lower temperatures. Polymer thick films will be covered in more detail in Sect. 30.3.4.

30.1 Thick Film Processing

Thick film technology is sometimes referred to as an *additive* technology in that the layers are built up in sequence without the need to remove (or subtract) parts of the film by techniques such as etching. Compare this with, say, a standard printed circuit board (PCB), where the conductive tracks are formed by selectively etching away the undesired areas (gaps) from a continuous copper layer. In the early days of the technology, the ability to fabricate together components made with different enabling technologies to produce a *hybrid* circuit was seen as a way of opening up a whole new field of electronics. This is still one of the most endearing features of the technology today.

30.1.1 Screen Printing

This is the distinguishing feature of thick film technology and is the method of depositing the desired films. The fundamental aspects of the process, traditionally recognized as an art, have been modernized and updated to provide a scientific tool for microelectronic technology. Figure 30.1 illustrates the main aspects of the screen printing process [30.3].

The screen fabric is permanently attached to the screen frame, which is firmly held within the screen printing machine. The distance between the screen and the substrate is typically around 0.5 mm, although the

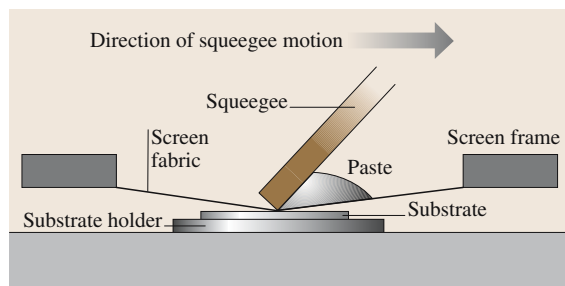


Fig. 30.1 The basic screen printing process

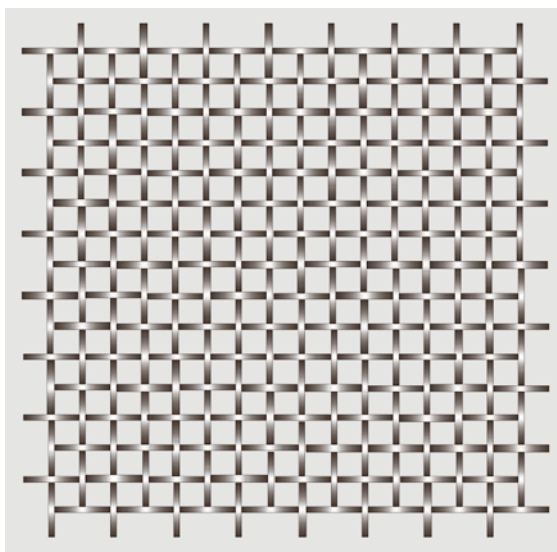


Fig. 30.2 A plain weave screen mesh

precise gap size is dependent on the overall screen dimensions. The substrate is held in place in the holder by either a vacuum or a mechanical clamp. The position of the screen can be finely adjusted to ensure good registration between consecutive layers. The paste is applied to the upper surface of the screen and the flexible *squeegee* is pulled across the screen over the ink, which is forced through the open areas of the screen mesh. At a point immediately behind the squeegee, the screen peels away from the substrate and, due to the surface tension between the ink and the substrate, leaves a deposit of the paste in the desired pattern on the substrate.

The squeegee is a flexible blade whose function is to transfer the ink through the screen and onto the substrate. It is usually made from materials such as polyurethane or neoprene. The squeegee pressure is adjustable and facilitates accurate and repeatable print thicknesses. The pastes generally exhibit pseudoplastic behavior in that

the viscosity varies with the applied shear force. This is a necessary property because the ink must have minimum viscosity to ensure good transfer through the screen, but it also must become more viscous after printing to provide a good definition of the film.

The function of the screen is to define the pattern of the printed film and also to control the amount of paste being deposited. The screens used in graphic artwork are generally made of silk and the process is often termed silk screen printing. For microelectronic circuits, however, the screens tend to be made from polyester, nylon or stainless steel. The resolution of the printed line widths is largely determined by the mesh count (in lines per inch or centimeter) and the mesh filament diameter. Nylon has very good elastic properties but relatively poor resilience. Stainless steel screens produce excellent line definition and durability and are particularly well suited to flat substrates. Polyester screens have the highest resilience and offer a long lifetime and low squeegee wear. Figure 30.2 shows a plain weave pattern on a typical thick film screen. An ultraviolet (UV)-sensitive emulsion covers the underside of the screen. The desired pattern is formed onto screen emulsion using a photographic process. The screen printing process can be used to deposit both cermet and polymer thick films; the processing steps that follow are, however, different.

30.1.2 The Drying and Firing Process

The final form of a cermet thick film is a fired, composite layer that is firmly attached to the substrate. Essentially, there are three main stages in the production process: screen printing, drying and firing. At each of these stages, the film is in a slightly different state. Commercial thick film pastes are purchased from the manufacturer in plastic jars and are similar to standard printing inks in many respects. The three main components in a thick film paste are:

- The active material;
- A glass frit;
- An organic vehicle.

The active material is a finely ground powder with a typical particle size of a few microns. A conductor paste, for example, will contain a precious metal or metal alloy. The glass frit serves as a binder that holds the active particles together and also bonds the film to the substrate. The organic vehicle is necessary to give the paste the correct viscosity for screen printing. It usually contains a resin dissolved in a solvent together with a surfactant that ensures good dispersion of the solid particles.

After screen printing, it is usual to allow the film to stand in air for a few minutes to let the paste level off. The film is then dried in an infrared belt drier or a conventional box oven at temperatures up to 150 °C. The purpose of drying is to remove the organic solvents so that the film and substrate can be handled during further processing steps. After drying, the films then proceed to the firing stage. In some circumstances, however, they may be overprinted with another thick film layer if the nature of the film allows this.

The high-temperature firing cycle performs three main functions: the remaining organic carrier is removed from the film, the electrical properties are developed, and the film is bonded to the substrate. In order to achieve these, it is necessary to subject the films to temperatures of up to 1000 °C in a moving belt furnace. Inside the furnace there are several heating zones, which can be independently controlled to a profile specified by the paste manufacturer. Figure 30.3 shows a typical firing profile that is used with commercial thick film resistor pastes. The substrates enter the furnace on a metal chain mesh belt at ambient temperature and slowly (at a rate of a few centimeters per minute) travel through the heating zones, which have their temperatures set to form the desired profile. The substrates remain at the peak temperature for about ten minutes and the electrical properties of the film are formed during this phase. Most thick film materials are fired in a clean, filtered air atmosphere, as this produces high-quality films with repeatable characteristics. Occasionally, however, it is necessary to provide an inert atmosphere such as nitrogen, so that materials such as copper can be processed.

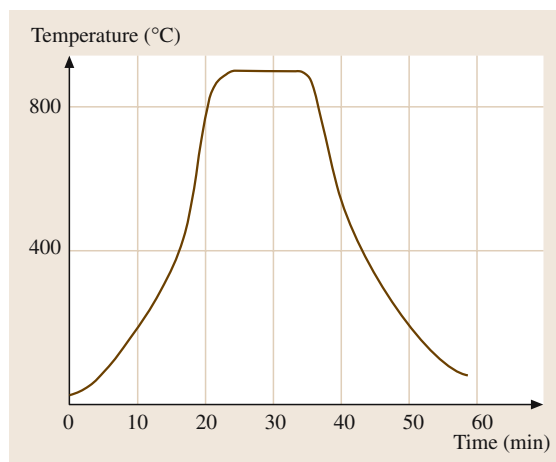


Fig. 30.3 A typical thick film firing profile with a peak temperature of 850 °C

30.2 Substrates

The main functions of the substrate are to provide mechanical support and electrical insulation for the thick films and hybrid circuits. Some of the main considerations for selecting substrates are listed below:

Dielectric constant: This determines the capacitance associated with different elements fabricated onto the substrate. The dielectric strength will also determine the breakdown properties of the substrate.

Thermal conductivity: Substrates with a high thermal conductivity can be used in applications where the circuit generates significant amounts of heat.

Thermal coefficient of expansion (TCE): In general terms, the TCE of the substrate should be closely matched to the thick film materials and other components mounted on it. In some cases this cannot be assured, and in this case the consequences (in terms of thermal strains) must be fully considered.

The main substrate materials used in thick film technology are the ceramic materials alumina (Al_2O_3), beryllia (BeO) and aluminium nitride (AlN). Enamelled or insulated stainless steel substrates are sometimes used in some applications. Silicon has also been used in specialist transducer applications. Alumina, however, is the most common substrate material and it possesses desirable physical and chemical properties in addition to providing an economical solution. Alumina of 96% purity is used in the vast majority of worldwide commercial circuits. The remaining 4% weight fraction of the content is made up of magnesia and silica, which improve the densification and electrical properties. Beryllia has a high thermal conductivity and is used in applications where rapid heat removal is required. It is, however, a very toxic material and is therefore only used in limited application areas. Aluminium nitride is, essentially, an alternative for beryllia, with a high thermal conductivity and also improved mechanical properties such as higher flexural strength. Insulated stainless steel substrates are sometimes used in applications where a high thermal dissipation and mechanical ruggedness are required. They are particularly well suited to mechanical sensor applications.

30.2.1 Alumina

Alumina substrates are manufactured by blending alumina powder, with an average particle size of around $1\text{ }\mu\text{m}$, together with small amounts of silica, magnesia and calcia. These are either ball- or roll-milled for about 10 h with lubricants, binders and solvents that en-

sure thorough mixing. Most thick film substrates are less than one millimeter thick, and the preferred method of fabrication is sheet casting. A slurry is allowed to flow out onto a smooth belt, and it passes under a metal doctor blade which controls the resultant thickness. The material is then dried in air to remove the solvent and, at this stage, it is sometimes referred to as the *green state* because of its color. The substrates are then fired in a kiln for at least 12 h. A peak temperature of around 1500°C ensures that the materials are properly sintered. During firing, the substrates can shrink by up to 20%, and this needs to be taken into consideration for the formation of the final substrate. The surface finish can be improved by coating the surface with a thin, glassy layer (glazing), which is done as an additional step at a lower temperature.

30.2.2 Stainless Steel

Stainless steel is a strong, elastic material with a relatively high thermal conductivity. Being a good electrical conductor, the steel must be coated with an insulating layer before it can be used as a substrate for thick film circuits. Porcelain enamelled steel substrates are made by coating a stainless steel plate with a glassy layer between 100 and $200\text{ }\mu\text{m}$ in thickness. The steel is enamelled either by dipping, electrostatic spraying or electrophoretic deposition of a low-alkali glass and subsequent firing at several hundred degrees Celsius. Some commercial paste manufacturers produce an insulating dielectric thick film ink that can be screen printed directly onto various types of stainless steel. The substrates are fired at a temperature of around 900°C . The paste contains a devitrifying glass that does not recrystallize on further firings and therefore provides compatibility with other standard thick film materials. Insulated stainless steel substrates also offer the advantages of having a built-in ground plane (the steel itself) and excellent electromagnetic and electrostatic shielding properties. It is also possible to machine the substrate, using conventional workshop facilities, prior to the circuit fabrication.

30.2.3 Polymer Substrates

In some applications it is desirable to fabricate a circuit onto a flexible substrate; evidence of these can be found in mobile telephones, calculators and notebook computers. Cermet thick film materials are not compat-

ible with flexible substrates and hence special polymer thick film materials are used (Sect. 30.3.4). Polymer substrate materials are mainly based on polyesters,

polycarbonates and polyimide plastics. The maximum processing temperature is usually limited to around 200 °C.

30.3 Thick Film Materials

30.3.1 Conductors

Thick film conductors are the most widely used material in thick film hybrid circuits. Their main function is to provide interconnection between the components in the circuit. For a multilayered circuit, the conductor tracks are separated by dielectric layers and connection between each layer is achieved with metallized vias. Conductors are also used to form attachment pads for surface-mounted components such as integrated circuits or discrete passive components (resistors, capacitors and inductors). They can also be used as bond pads for naked dice, which may be attached directly to the thick film circuit. Another function of conductors is to provide the terminations for thick film resistors. With such a diverse range of applications, it is no surprise that a wide range of conductor materials is available.

The characteristics of thick film conductors are dependent upon the composition of the functional phase of the paste. Typically, these comprise finely divided particles of precious metals such as silver, gold, platinum or palladium. Base metals such as aluminium, copper, nickel, chromium, tungsten or molybdenum are also used. The particle size, distribution and shape also have an effect on the electrical and physical properties of the fired film.

The resistance of a conductor film is given by

$$R = \frac{\rho l}{wt},$$

where ρ is the bulk resistivity ($\Omega \text{ cm}$), l and w are the length and width of the conductor respectively, and t

is the fired thickness of the film. The term *sheet resistivity* is often used for thick films and is defined as the bulk resistivity at a given thickness (ρ/t), expressed in ohms per square (Ω/\square). This is convenient because the resistance of the track can then be calculated by simply multiplying the sheet resistivity by the aspect ratio (l/w) of the film. Table 30.1 summarizes some of the most common metals and metal alloys used in thick film conductors together with their sheet resistivities.

Silver Conductors

Silver pastes were one of the earliest thick film conductors to be developed. They possess good bond strength and high conductivity. There are, however, several disadvantages to using pure silver which prevent it being widely used in many applications. These include:

- Poor leach resistance to solder;
- Oxidation in air over time;
- Susceptible to electromigration in the presence of moisture, elevated temperature and bias voltage.

Silver/palladium conductors

The alloy of silver and palladium is the most common type of thick film conductor and is the alloy most widely used in the hybrid circuit industry. It can be used for interconnecting tracks, attachment pads and resistor terminations, but is generally not recommended for wiring bonding pads. Silver/palladium conductors overcome many of the problems associated with pure silver, and low-migration formulations are available from several commercial suppliers.

Gold Conductors

Gold pastes have a high conductivity and are mainly used in applications where high reliability is required. Gold is a particularly good material for wire bonding pads, although it has relatively poor solderability. Gold is a precious material and hence very expensive; it is therefore not used for general purpose applications and is limited to those areas that can justify the higher costs.

Table 30.1 Most of the common metals and metal alloys used in thick film conductors and their sheet resistivities

Metallurgy	Sheet resistivity ($\text{m}\Omega/\square$)
Silver (Ag)	1–3
Gold (Au)	3–5
Copper (Cu)	2–3
Silver/palladium (Ag/Pd)	10–50
Gold/palladium (Au/Pd)	10–80
Gold/platinum (Au/Pt)	50–100

Copper Conductors

Copper thick films have to be processed in an inert atmosphere; nitrogen is typically used in this case. Copper is widely used as the conductor material for printed circuit boards owing to its good electrical conductivity and solderability. These features are also applicable to thick film copper conductors, which also offer ability to handle larger currents than most other thick film conductors.

Platinum Conductors

Not too surprisingly, perhaps, these are the most expensive of all the commercial thick film conductor materials. Platinum films have a very high resistance to solder leaching and exhibit similar electrical properties to those of the bulk material: a linear, well-defined temperature coefficient of resistance (TCR). Platinum films are therefore used in specialist applications such as heaters, temperature sensors and screen-printed chemical sensors.

Gold Alloy Conductors (Au/Pd, Au/Pt)

These have good bond strength, good solder leach resistance and are relatively easy to solder. The conductivity of the gold alloys tends to be inferior to that of the other type of thick film conductor, but they can be used with both ultrasonic (aluminium) and thermosonic (gold) wire bonds.

Once the conductor has been fired, a composite structure is formed. Figure 30.4 shows a simplified view of the cross-section of a fired thick film conductor on a substrate. Many of the metallic particles have joined together to make a continuous chain. The glass is mainly evident at the interface between the bulk of the film and the substrate. The presence of voids (gaps), both within the film and on its surface, can also be noted.

30.3.2 Resistors

Resistor inks consist of a mixture of the three main phases listed earlier. The relative proportions of the active material to glass frit can have a dramatic effect on the electrical properties of the fired film. The earliest resistors were made from materials such as carbon, silver and iron oxide and were found to suffer from poor long-term stability, unpredictability of fired resistivity, and unacceptably high temperature coefficients. Modern thick film resistor systems are mainly based on ruthenium dioxide (RuO_2) and have much-improved characteristics. This material has a high conductivity and is extremely stable at high temperatures. Within the resistive paste, the conductive phase comprises submicron

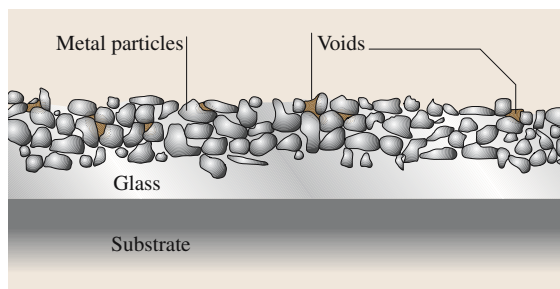


Fig. 30.4 Idealized cross-sectional view of a cermet thick film conductor

particles, and these are mixed with larger glass particles (several microns in diameter). Various additives are added to the formulation to improve the stability and electrical properties of the fired film. The nature of these additives is proprietary knowledge of the paste manufacturer.

Figure 30.5 shows an idealized cross-section of a thick film resistor fabricated onto a substrate. The end terminations are first printed and fired and then the resistor is formed in the same manner, allowing for a slight overlap on the terminations so that small misalignments during the printing of the resistor can be accounted for. The actual surface profile of a real resistor would show a nonuniform thickness across its length.

Commercial thick film resistor pastes are available in a range of sheet resistivities from 1 to $10^9 \Omega/\square$, and the value of the fired resistor is determined by the selected sheet resistivity and the ratio of the length to the width, as previously described for conductors. For example, a $10 \text{ k}\Omega$ resistor could be made by printing a square ($l = w$) of $10 \text{ k}\Omega/\square$ resistor paste. Similarly, a $5 \text{ k}\Omega$ resistor requires an aspect ratio of 1 : 2 ($l : w$) if printed with the same paste. Note that the absolute value of a thick film resistor is not limited to a preferred value, as is the case for discrete resistors.

Figure 30.6 shows the temperature dependence of resistance for a typical thick film resistor. The shape of

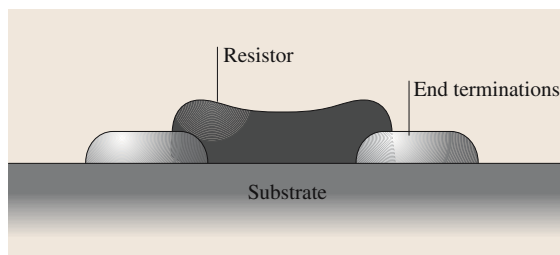


Fig. 30.5 An idealized cross-section of a thick film resistor

the plot is unusual because there is a point at which the resistance is a minimum (T_{\min}). In general terms, the sensitivity of a resistor to temperature is denoted by the temperature coefficient of resistance (TCR). Mathematically, this can be expressed by

$$\text{TCR} = \frac{\Delta R/R}{\Delta T},$$

where $\Delta R/R$ is the relative change in resistance and ΔT is a small change in temperature. Metals generally exhibit an increase in resistance with increasing temperature, so the TCR is positive. A thick film resistor appears to have both a positive and negative TCR at different regions of the resistance versus temperature curve [30.4]. The point of minimum resistance occurs around room temperature, which means that the resistance is stable under normal operating conditions. It is common practice to refer to the 'hot' and 'cold' TCR for the regions above and below T_{\min} , respectively.

The long-term stability of thick film resistors in air at room temperature is excellent, with a typical change of less than 0.2% over 10^5 h (about 100 years!). At elevated temperatures the stability is still impressive, with most resistance values changing by only 0.5% over 1000 h at 150 °C. The tolerance of fired resistors is around $\pm 20\%$ due to a variety of reasons, including variations in resistor thickness, the tolerance on the quoted sheet resistivity and the effects of the firing process. If accurate values of resistance are required, it is therefore necessary to trim the resistor. This is achieved by removing areas of the resistor with either an air-abrasive jet or a laser. Resistor values can only be increased by trimming, so the printed value is always designed to be less than that of the post-trimmed value.

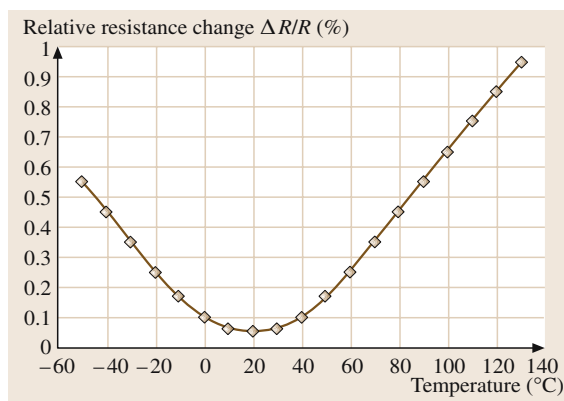


Fig. 30.6 A plot of relative resistance change against temperature for a thick film resistor

30.3.3 Dielectrics

Dielectric pastes have four main uses in thick film hybrid circuit applications:

- Cross-over insulators for multilayer circuits;
- Thick film capacitor dielectrics;
- Passivation layers;
- Insulation layers for stainless steel substrates.

Cross-over dielectrics comprise a ceramic material such as alumina, together with the usual glass frit and organic vehicle. These films are required to have a low dielectric constant in order to minimize capacitive coupling between the conducting tracks. They must also have a good insulation resistance and a smooth, pinhole-free surface finish.

There is an occasional requirement to fabricate thick film capacitors, although the wide availability of surface-mounted capacitors has largely precluded their need. It is also difficult to ensure that post-fired values of thick film capacitors closely match those for which they were designed. An expensive trimming process is needed to ensure that the correct values are obtained.

Overglazes are mainly used to protect resistors from overspray during trimming and also the circuit from environmental attack during operation. The passivation layer is usually the last one to be processed and it is therefore fired at a reduced temperature in order to minimize any adverse refiring effects on other layers. Overglaze materials are therefore almost exclusively made from low-temperature glasses that fire at temperatures of 450 to 500 °C.

The use of insulated stainless steel substrates was mentioned earlier. The dielectric materials used for this purpose must provide a high insulation resistance (in excess of $10^{12} \Omega$) and also possess a high breakdown voltage (greater than 2 kV/mm). There is an inherent mismatch between the thermal coefficient of expansion of the metal substrate and that of the insulation layer, which limits the type of steels that can be used for this purpose.

30.3.4 Polymer Thick Films

The processing temperatures for polymer thick films are significantly less than those needed for cermet materials. Rather than being 'fired', polymer materials are cured at temperatures below 200 °C. One advantage of polymer thick films over cermets is a reduction in processing and material costs. As with cermet pastes, the formulation of polymer thick films comprises the active

material, a polymer matrix and various solvents. The polymer matrix acts as ‘glue’ for the active component. Three types of polymer organic composition are used in polymer thick films:

- Thermoplastic;
- Thermosetting;
- Ultraviolet (UV)-curable.

With thermoplastics, the required viscosity for screen printing is achieved via solvents. The polymer material is typically acrylic, polyester, urethane or vinyl. After printing, the paste is hardened by drying in a belt or box oven. These types of film have relatively poor resistance to environmental conditions and are not resistant to elevated temperatures and solvents. Thermosetting pastes have polymers that are partially cured and are typically epoxy, silicone or phenolic resin. After printing, the polymer is fully cured, providing a strong and stable matrix. Solvents are still needed to provide the correct rheology for printing. The UV curable pastes are generally used for dielectric inks and can be cured at room temperature under an ultraviolet light source.

For conductors, the most commonly used active phases are silver, copper and nickel. Owing to their poor stability at high temperatures, polymer thick film conductors cannot be soldered and alternative forms of attachment must be adopted. Carbon is typically used as the active material in polymer thick film resistors. The performance of these resistors is inferior to that of their cermet counterparts and they are therefore seldom used in critical applications. The dielectric pastes are similar in nature to conductors and resistors except that the conducting phase is omitted from the formulation. Some manufacturers add minerals to improve the electrical and mechanical properties of the dielectric films. Polymer thick films are a popular choice of material for disposable biosensors such as those used in

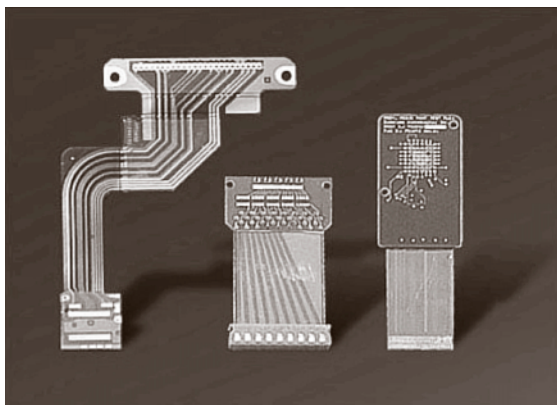


Fig. 30.7 A selection of flexible circuits (courtesy of Flex Interconnect Technologies, Milpitas, USA)

the home testing of levels of glucose in human blood samples.

The process for fabricating polymer thick films is similar to that used with cermet materials. Once they have been screen printed, the layers are left to stand in air for a few minutes to ensure that the surface is level and contains no residual mesh patterns. The curing process is achieved in a box oven or an infrared belt dryer at temperatures in the range 150–200 °C. For a thermoset polymer, the higher the temperature and longer the curing time, the greater the cross-linking of the polymer chains in the matrix. This can lead to improved film stability and increased shrinkage.

An early and successful application of polymer thick film technology was the fabrication of membrane switches for keyboards. Today, examples of polymer thick film circuits can be found in many consumer products, such as mobile phones, portable computers, personal digital assistants and calculators. Figure 30.7 shows some typical flexible polymer circuits.

30.4 Components and Assembly

30.4.1 Passive Components

Passive electronic components are those that do not require an external energy source to function. Examples are resistors, capacitors and inductors. As we have seen earlier, thick film technology allows the fabrication of high-quality, stable resistors. It is also possible to add resistors to a thick film circuit in the form of an additional chip component. Such devices are

available as surface-mounted devices, which do not require holes to be drilled into the circuit board. Interestingly, chip resistors are often manufactured as multiple parts using thick film techniques on ceramic substrates. These are then diced and the terminations are added. An example of a typical chip component is shown in Fig. 30.8. Resistor values can range 1 Ω to 10 MΩ, with typical tolerances of between ±1% to ±20%.

Planar screen-printed capacitors are rarely used in thick film hybrid circuits owing to their poor stability and high production costs. They typically comprise at least three layers (two electrodes and one dielectric layer), and trimming is often required to achieve the target value of capacitance. Chip capacitors offer improved performance at a lower cost. Capacitances in the range 1 pF to 100 μ F are readily available in a range of sizes from the 0201 series (0.6 mm \times 0.3 mm \times 0.3 mm) to the 2220 series (5.7 mm \times 5.0 mm \times 3.2 mm). Chip inductors are also available (typical range: 0.1 μ H to 1000 μ H) in a variety of package sizes. There is also a wide choice of variable passive components such as potentiometers and variable capacitors/inductors that are currently available from many major component suppliers.

30.4.2 Active Components

Active components are those that require an external energy source to function. Examples are transistors, diodes

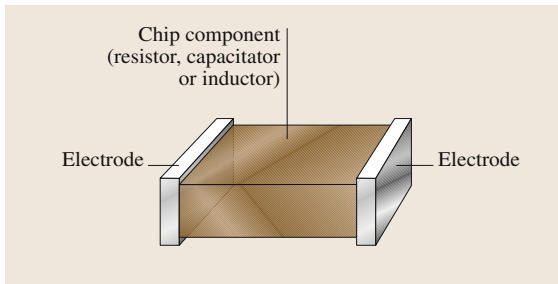


Fig. 30.8 A typical surface-mounted chip component

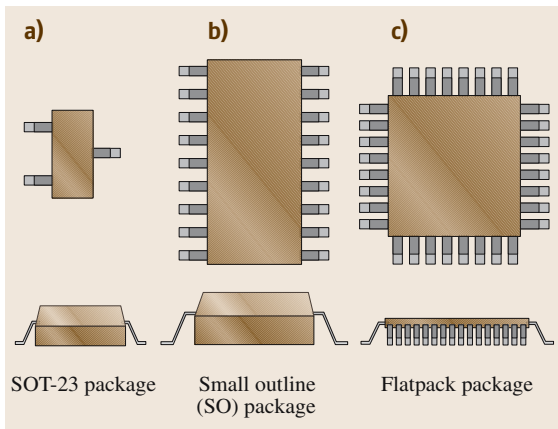


Fig. 30.9a–c Examples of surface-mounted packages (a) SOT-23 package (b) Small outline (SO) package (c) Flatpack package

and semiconductor integrated circuits. A wide variety of semiconductor components are available to the thick film circuit designer. Transistors and diodes, being relatively small devices, are obtained in a small plastic package with three terminals. This is known as a SOT-23 package and is shown in Fig. 30.9a. Many standard integrated circuits that are available in dual-in-line (DIL) packages for through-hole printed circuit boards are also available in small outline (SO) packages for hybrid circuits. An example of a small outline device is shown in Fig. 30.9b, and these usually have between 8 and 40 pins. For devices with higher pin counts such as microprocessor, gate arrays and so on, it is usual to place the leads on all four sides of the package, as with the flatpack device shown in Fig. 30.9c. Occasionally, ICs may be obtained in the form of a naked die, without any packaging. In such cases, it is necessary to glue the chip to the board and to bond very fine wires from the chip to the board. Care has to be taken to ensure that the naked device is suitably encapsulated for use afterwards.

Other forms of IC include flip chips, which are essentially naked chips with raised connection contacts (bumps) made of solder, gold or aluminium. These are mounted by turning over the chip (flipping) and bonding directly to the substrate. Beam lead chips also examples of naked dice with either gold or aluminium leads protruding from the edge. The leads (beams) are an integral part of the chip metallization process. Such devices are usually passivated with a layer of silicon nitride during processing. Tab automated bonding (TAB) refers to a technique by which the naked chip is attached to metallized fingers on a continuous strip of film. Large quantities of devices can be produced on a single roll and the leads of the devices are welded onto the boards before the carrier film is removed. In this manner, it is possible to achieve good yields on high-density circuit populations.

30.4.3 Trimming

The tolerance on the printed value of thick film components such as resistors is around $\pm 20\%$ of the desired value. Many applications require a much tighter tolerance, and so the components need to be trimmed. The two most popular techniques used are trimming by laser or by an air-abrasive jet. Both of these are capable of producing resistors with a tolerance of $\pm 0.1\%$. Laser trimming, however, is more amenable to large-scale component adjustment.

Air-abrasive trimming uses a pressurized jet of air containing a fine abrasive powder to remove a small

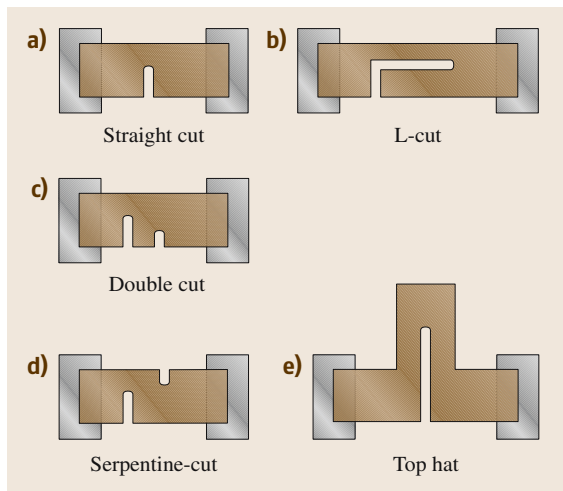


Fig. 30.10a–e Examples of trim cuts for resistors (a) Straight cut (b) L-cut (c) Double cut (d) Serpentine cut (e) Top hat

area of the fired thick film. The diameter of the jet nozzle is around 0.5 to 1 mm. Alumina particles of average diameter 25 μm are often used as the abrasive medium. The substrate containing the component to be trimmed is held underneath the nozzle at a distance of around 0.6 mm and electrical probes are attached to the device. During trimming, the particles in the jet stream remove material from the component. The debris is removed from the substrate by a vacuum exhaust system. In the case of resistor trimming, the value can only increase because the material is being removed. Resistor values cannot be reduced by the trimming process. For this reason, resistors requiring trimming are designed to be between 25 and 30% lower than the post-trimmed value.

Laser trimming has the advantage of offering a fully automated, high-speed way of adjusting component values. The thick film is vaporized with high-energy laser beam pulses. The laser is typically a Q-switched neodymium-doped YAG (yttrium aluminium garnet) type. A single pulse removes a hole of material and a line is achieved by overlapping consecutive pulses.

Resistors are the most common component requiring trimming. Sometimes conductor tracks need to be adjusted in special circumstances. For example, a platinum conductor being used as a classic Pt 100 resistance thermometer must have a resistance of 100 Ω at 0 $^{\circ}\text{C}$. In rare circumstances, thick film capacitors can be trimmed by removing an area of one of the plates, although the post-trimmed stability is poor.

Figure 30.10 shows some examples of different types of cut that can be used to trim resistors. The straight cut is the fastest way to trim, but it does not provide a reliable way of achieving a high accuracy. The L-cut overcomes this problem; the resistor is first trimmed straight and then the cut runs parallel to its length, providing a finer adjustment of the value. Both the serpentine and double cut require additional cuts perpendicular to an initial straight cut. The top hat structure is used in situations where a large change in resistance is needed.

30.4.4 Wire Bonding

When a naked integrated circuit is needed as part of a hybrid circuit, connection must be made directly to the bond pads on the die. The chip can be attached to the substrate using an epoxy or by a gold/silicon eutectic bond. Once the chip is firmly held in position, wire bonding can commence. Three main methods are:

- Thermocompression;
- Ultrasonic;
- Thermosonic.

Thermocompression bonding relies on a combination of heat and pressure. The wire is usually made of gold with a diameter of around 25 μm . Gold or palladium/gold pads are deposited onto the hybrid substrate prior to bonding. The wire is fed through a ceramic capillary and a ball is formed at the end of the wire by a flame or spark discharge. A temperature of 350 $^{\circ}\text{C}$ is required for the bond and is achieved by heating either the substrate or the capillary. An epoxy chip bond cannot be used with this technique as it will soften during the bonding process. The first part of the bond is made on the aluminium bond pad on the chip; the capillary is lowered onto the pad and a force is applied to form a ball shape on the pad. The wire feeds out of the capillary and is then positioned over the desired pad on the substrate. As the capillary is lowered, a combination of heat and pressure forms the bond and the wire is then broken so that the process can be repeated for further bonds.

Ultrasonic bonding uses either a gold or aluminium wire and does not require external heat. The ultrasonic energy is supplied from a 40 kHz transducer. The combination of pressure and ultrasonic vibration causes the materials to bond together at the interface of the wire and the bond pad. This is generally a faster technique than the thermocompression method.

The final category, thermosonic bonding, is a combination of the other techniques. The substrate is heated to around 150 $^{\circ}\text{C}$ and the bond is made using the ultra-

sonic vibrations. This method is amenable to multilevel and multidirectional bonding and is therefore the preferred method, allowing bonding of up to 100 wires per minute.

30.4.5 Soldering of Surface-Mounted Components

Several techniques exist for the attachment of surface-mounted components to a thick film hybrid circuit. Surface-mounted components are generally much smaller than their through-hole counterparts. Of course, soldering by hand is also possible, although this is a tricky task requiring good operator skill and is often impractical because of the relatively long length of time needed.

Solder dipping requires the components to be placed on the board, either by hand or by a special pick-and-place machine. The components are fixed in position on the substrate by adding a small dot of glue and elevating the temperature to between 120 °C and 180 °C, which is sufficient to cure the adhesive. The board can then be dipped into a bath of molten solder at a temperature of 200 °C and then withdrawn at a sufficient rate to ensure that an adequate solder coating is obtained.

Wave soldering also requires the components to be fixed in position prior to the soldering process. Wave soldering machines were originally used for soldering through-hole components onto printed circuit boards, but they can also be used effectively with surface-mounted devices. The substrates are placed on a moving belt component side-down and initially pass through a flux bath before entering a solder bath. A wave of molten solder then flows over the substrate and creates a good joint at the desired location. This process can expose the components to a great thermal shock and it is therefore common to have a preheating phase which minimizes such effects. With both of these techniques, it is also necessary to ensure that a solder resist layer is applied to the substrate to cover all the areas that are not required to be soldered. With very densely populated circuits, there can also be a masking effect where some areas are not sufficiently coated with solder.

Reflow soldering is the preferred method of attaching surface-mounted devices. A solder cream is deposited onto the component pads either by screen printing or by a solder dispenser. The flux within the cream is sufficiently tacky to hold the component in place so that handling of the substrate is possible. After all of the components have been positioned on the circuit, the solder cream is dried and then reflowed.

This process takes place by belt reflow, vapor phase or infrared belt system. A typical belt reflow system comprises a thermally conducting belt upon which the substrates are placed. The belt then travels through a number of heating stages, which causes the solder cream to melt (reflow). Vapor-phase soldering requires the substrates to be lowered into a vessel containing a boiling, inert fluorocarbon. The vapor condenses onto the substrate and raises the temperature uniformly to that of the liquid below. Infrared belt reflow systems are similar to those used for drying thick film materials. The substrate is placed on a wire-mesh belt, which travels through several infrared radiator zones.

30.4.6 Packaging and Testing

Thick film hybrid circuits are very versatile and offer advantages over other forms of enabling technologies. Owing to this flexibility, the circuits have a wide range of shapes and sizes and hence there is no “standard” package type. Selection of a particular form of packaging must therefore involve the consideration of issues such as:

- Protection of the circuit from harsh environmental conditions;
- Protection from mechanical damage;
- Avoidance of water ingress;
- Electrical or mechanical connections to other parts of the system;
- Thermal mismatches of different materials.

A simple way of protecting the circuit is to screen print an overglaze layer over the substrate, covering all areas of the substrate except those where components are to be added. A lead frame can then be added to the substrate to allow external connections to be made. An example of a thick film hybrid circuit (without overglaze) is depicted

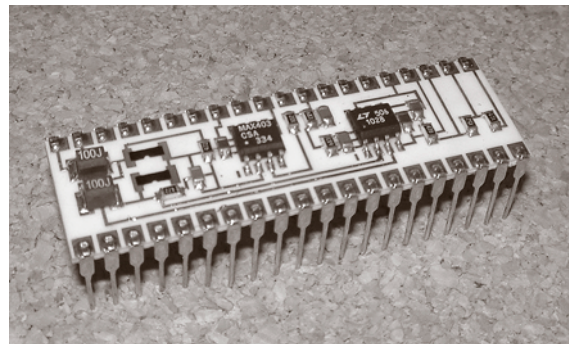


Fig. 30.11 A thick film hybrid circuit

in Fig. 30.11. The two resistors on the left of the circuit have been trimmed and the straight cuts are visible.

Conformal coatings are often used to protect the circuit from environmental attack. These are applied in the form of either a powder or fluid. In the former case, the substrate is heated and immersed into the powder. The temperature is then increased so that the coating dries. For fluids, the substrate is dipped into the coating material and subsequently dried at a temperature of around 70 °C. Typically, the thickness of a conformal coating is between 300 and 1200 μm.

For circuits requiring operation in harsh environments, a special hermetic packaging is needed. The package can be made from ceramics, metals, ceramic/metal or glass/ceramic compositions. The hermetic seal is made by brazing, welding or glass sealing

techniques. This form of packaging is often very expensive and is therefore only used in special application areas.

The final stage of the process is to test the circuit to see that its performance matches the design specification. Electrical testing can be difficult if the circuit has been coated or hermetically sealed, as physical access to components may be restricted. It is therefore usual to ensure that key test points are brought out to an external pin on the package. Environmental testing over a range of temperature and humidity may also be required in some circumstances. High-reliability circuits are often subject to a so-called burn-in phase, which involves holding the circuits at an elevated temperature for a given time to simulate the ageing process.

30.5 Sensors

Advances in the field of sensor development are greatly affected by the technologies that are used for their fabrication. The use of thick film processes as an enabling technology for modern-day sensors continues to expand. As we have already seen, the ability to produce miniaturized circuits is clearly one area in which thick film technology excels. The hybrid electronic circuitry can be integrated into the sensor housing to produce the basis of a smart (or intelligent) sensor [30.5]. Thick film technology also offers the advantage that it can provide a supporting structure onto which other materials can be deposited, possibly using other enabling technologies [30.6].

A major contribution of the technology to sensor development, however, results from the fact that the thick film itself can act as a primary sensing element. As an example, the thick film strain gauge, described below, is merely a conventional thick film resistor that is configured in such a way as to exploit one of its physical characteristics. Commercial thick film platinum conductors can be trimmed and used as calibrated temperature sensors. Most standard pastes, however, have not been specifically developed for sensor applications and do not necessarily have optimum sensing properties. The formulation of special-purpose thick film sensor pastes is the subject of intensive research activity [30.7].

For the purpose of this text, a *sensor* is considered as being a device that translates a signal from one of the common sensing domains (mechanical, thermal, op-

tical, chemical or magnetic) into an electrical signal. An *actuator* is a device that converts an electrical signal into one of the other domains (mainly mechanical).

30.5.1 Mechanical

In broad terms, thick film mechanical sensors are mainly based on piezoresistive, piezoelectric or capacitive techniques. Materials that exhibit a change in bulk resistivity when subjected to deformation by an external force are termed *piezoresistive*. A more common term is the strain gauge, denoting the fact that such devices produce a change in resistance when strained. The effect can be observed in standard cermet thick film resistors [30.6,8]. The sensitivity of a strain gauge is called the gauge factor and is defined as:

$$GF = \frac{\Delta R/R}{\varepsilon},$$

where $\Delta R/R$ is the relative change in resistance and ε is the applied strain (dimensionless). The gauge factors of metal foil strain gauges and thick film resistors are around 2 and 10 respectively. The former have typical resistance values of either 120 Ω or 350 Ω. As we have already seen, however, it is possible to produce thick film resistors with a wide range of resistance values, and this allows greater flexibility in strain gauge design. It is usual to place the strain gauges in a Wheatstone bridge configuration in order to produce a linear output analog voltage change that is proportional to the mechanical

measurand. A wide range of thick film piezoresistive sensors exist, including accelerometers, pressure sensors and load cells.

Piezoelectric materials exhibit the property of producing an electric charge when subjected to an applied mechanical force. They also deform in response to an externally applied electric field. This is an unusual effect, as the material can act as both a sensor and actuator. Certain crystals such as quartz and Rochelle salt are naturally occurring piezoelectrics, whilst others, like the ceramic materials barium titanate, lead zirconate titanate (PZT) and the polymer material polyvinylidene fluoride (PVDF), are *ferroelectric*. Ferroelectric materials are those that exhibit spontaneous polarization upon the application of an applied electric field. This means that ferroelectrics must be poled (polarized) prior to use in order to obtain piezoelectric behavior.

Thick film piezoelectrics have been made by mixing together PZT powder, a glass binder and an organic carrier [30.9]. A conducting layer is first screen printed, dried and fired onto a substrate and then several layers of the PZT film are deposited onto this lower electrode. The piezoelectric layer can be processed in a similar manner to conventional thick films. An upper electrode layer is then deposited onto the PZT in order to make a sandwich structure similar to that shown in Fig. 30.12. This is essentially a planar capacitor and the as-fired film must undergo a poling process by applying a DC electric field of around 4 MV/m and elevating the temperature to about 120 °C.

Thick film piezoelectrics have been used in a variety of sensor and actuator applications, including accelerometers, pressure sensors [30.10], micromachined pumps [30.11], surface acoustic wave (SAW) [30.12] and resonant sensors [30.13].

The capacitance C of a parallel plate capacitor is given by

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d}$$

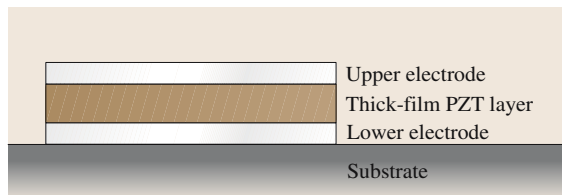


Fig. 30.12 Cross-section of a thick film piezoelectric sample

where ε_0 is the permittivity of free space, ε_r is the relative permittivity of the material between the electrodes, A is the area of overlap, and d is the separation of the electrodes. A mechanical sensor exhibiting a change in capacitance can be made by varying A , d or by displacing the dielectric (changing ε_r). The most popular configuration is to vary d in accordance with the desired measurand. This results in a nonlinear relationship between displacement and capacitance. If the variable plate is positioned between two fixed outer electrodes, then a differential structure with a linear response can be obtained. This is a common arrangement in many types of pressure sensor.

30.5.2 Thermal

Devices that exhibit a change in resistance in accordance with variations in temperature are termed *thermoresistive*. For metals, such devices have a linear response to temperature and are known as resistance thermometers. Thermally sensitive semiconductors, typically having a nonlinear response, are termed *thermistors*. Thick film platinum conductor layers can be used as resistance thermometers and exhibit a linear TCR of around 3800 ppm/°C, slightly lower than that of bulk platinum, which is around 4000 ppm/°C. Platinum resistance thermometers (PRTs) are often trimmed so that they have a resistance (R_0) of 100 Ω at 0 °C. Such sensors are sometimes referred to as Pt 100s and are often made of wound platinum wire. Thick film Pt 100s, fabricated onto alumina substrates, are available commercially and are considerably cheaper than the bulk versions. The following expression applies to a PRT over the linear part of the temperature characteristic (−200 °C to +500 °C)

$$R = R_0(1 + \alpha T),$$

where R is the resistance at a temperature T and α is the TCR (ppm/°C).

Thermistors are usually available in the form of discs, rods or beads comprising a sintered composite of a ceramic and a metallic oxide (typically manganese, copper or iron). Thermistor pastes are commercially available and screen-printed sensors are fabricated in a similar manner to conventional thick film resistors. The electrodes are first printed and fired onto an alumina substrate. The thermistor is then deposited across the electrodes and can be trimmed to a specific value if desired. Most thermistors have a negative temperature coefficient (NTC) of resistance; in other words their resistance decreases as the temperature increases. The

resistance versus temperature relationship is of the form

$$R = R_0 \exp \left[\beta \left(\frac{1}{T} - \frac{1}{T_0} \right) \right]$$

where R_0 is the resistance at a reference temperature T_0 (usually 25 °C).

Positive temperature coefficient (PTC) thermistors are also available, although they are generally not as stable or repeatable as NTCs and are therefore used as simple thermal detectors rather than calibrated devices.

Another type of temperature sensor can be made by joining together two dissimilar metals (or semiconductors). If a temperature difference exists between the joined and open ends, then an open-circuit voltage can be measured between the open ends. This is known as the Seebeck effect and is the basis of a *thermocouple*. A thick film version can be made by overlapping different conductor materials on a substrate, although the thermal sensitivity is much less than traditional wire-based devices.

30.5.3 Optical

Screen-printed photosensors are probably one of the earliest examples of thick film sensors, and their use dates back to the mid 1950s. Materials that exhibit a change in electrical conductivity due to absorbed electromagnetic radiation are known as *photoconductors*. Cadmium sulfide (CdS) is an example of such a material and is notable because of its highly sensitive response in the visible range (450–700 nm). The resistances of such devices can drop from several tens of MΩ in the dark to a few tens of ohms in bright sunlight.

Thick film photoconductor pastes are not widely available commercially, but have been the subject of some research activity [30.14]. Such pastes, based on cadmium sulfide and selenide, are prepared by sintering powdered CdS or CdSe with a small amount of cadmium chloride (which acts as a flux) at a temperature of around 600 °C. This is then ground into a powder and mixed with an organic carrier to make a screen-printable paste. This can then be printed over metal electrodes and fired at a temperature of around 600 °C in air.

30.5.4 Chemical

Thick film materials have been used in a variety of chemical sensor applications for the measurement of gas and liquid composition, acidity and humidity [30.15]. The two main techniques are impedance-based sensors and electrochemical sensors. With the former method, the

measurand causes a variation in resistance or capacitance, whilst the latter relies on the sensed quantity changing an electrochemical potential or current.

Impedance-based gas sensor pastes usually comprise a semiconducting metal oxide powder, inorganic additives and organic binders [30.16]. The paste is printed over metal electrodes and a back-heated resistor on an alumina substrate. The heating element is necessary to promote the reaction between the gas being measured and the sensing layer. Figure 30.13 shows an example of a thick film sensor, without a heating element, that can be used to measure humidity. The interlocking ‘finger’ electrodes are often referred to as interdigitated electrodes and are screen printed and fired onto an alumina substrate. A porous dielectric layer is then screen printed onto the electrodes. As the humidity increases, moisture will penetrate the surface of the dielectric layer causing a change in dielectric constant within the sensitive layer. This results in a change in capacitance between the electrodes.

Electrochemical techniques can be used to realise pH sensors. These are often used in biomedical, fermentation, process control and environmental applications. These devices often make use of a solid electrolyte which generates an electrochemical potential between two electrodes in response to the measurand.

Perhaps the most common example of a thick film chemical sensor is the disposable, polymer-based glucose sensor used in many home testing kits for diabetic patients. This illustrates how thick film sensors can offer robust, compact and cost-effective solutions to many modern-day requirements.

30.5.5 Magnetic

Some screen-printable conductors, particularly those containing nickel, exhibit a change in resistivity in response to an applied magnetic field. Such devices are referred to as *magnetoresistive* sensors. Air-fireable

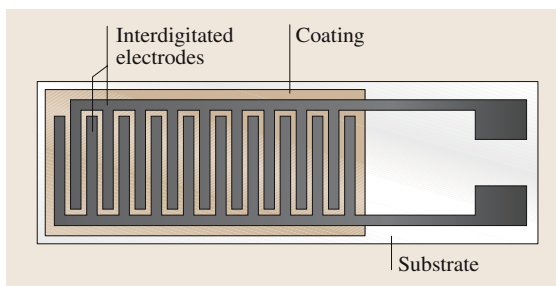


Fig. 30.13 A thick film humidity sensor

nickel-based conductors have been shown to exhibit a nonlinear change in resistance for a linear increase in applied magnetic field [30.17]. A peak change in resistance of around 1% can occur at an applied field of 0.1 T. Researchers have made linear and rotary displacement sensors based on thick film nickel pastes, although it should be noted that such devices are also thermoresistive and therefore the magnetic measurements need to be taken in a temperature-controlled environment.

30.5.6 Actuators

We have previously defined an actuator as a device that converts a signal from the electrical domain into one of the other signal domains. It was noted earlier that piezoelectric materials produce a mechanical stress in

response to an electrical charge. Such materials can therefore be used as actuators. Thick film piezoelectric layers have been screen printed onto thin silicon diaphragms in order to form the basis of a micropump [30.11].

Photovoltaic devices convert incident optical radiation into electric current and are often termed *solar cells*. They are used to power devices such as calculators, clocks, pumps and lighting. In general terms, the output power level is proportional to the physical size of the photovoltaic cell. The device is essentially a heterojunction between n-type and p-type semiconductors. Thick film solar cells have been made comprising CdS (n-type) and CdTe (p-type) as the junction materials [30.18]. Such thick film actuators have been shown to have relatively low conversion efficiencies (between 1% and 10%).

References

- 30.1 R. A. Rikoski: *Hybrid Microelectronic Circuits: The Thick-Film* (Wiley, New York 1973)
- 30.2 M. A. Topfer: *Thick-Film Microelectronics: Fabrication, Design and Fabrication* (Van Nostrand-Reinhold, New York 1971)
- 30.3 P. J. Holmes, R. G. Loasby: *Handbook of Thick Film Technology* (Electrochemical Publ., Ayr 1976)
- 30.4 M. Prudenziati, A. Rizzi, P. Davioli, A. Mattei: *Nuovo Cim.* **3**, 697–710 (1983)
- 30.5 J. E. Brignell: *Thick-Film Sensors*, ed. by M. Prudenziati (Elsevier, Amsterdam 1994)
- 30.6 J. E. Brignell, N. M. White, A. W. J. Cranny: Sensor applications of thick-film technology, *IEE Proceedings Part I, Solid State and Electron Devices* **135**(4), 77–84 (1988)
- 30.7 N. M. White, J. D. Turner: *Meas. Sci. Technol.* **8**, 1–20 (1997)
- 30.8 C. Canali, D. Malavisi, B. Morten, M. Prudenziati: *J. Appl. Phys.* **51**, 3282–3286 (1980)
- 30.9 H. Baudry: Screen printing piezoelectric devices, 6th European Microelectronics, 456–463 (Bournemouth, UK, 1987)
- 30.10 M. Prudenziati, B. Morten, G. De Cicco: *Microelectron. Int.* **38**, 5–11 (1995)
- 30.11 M. Koch, N. Harris, A. G. R. Evans, N. M. White, A. Brunnschweiler: *Sensors Actuat. A* **70**(1–2), 98–103 (1998)
- 30.12 N. M. White, V. T. K. Ko: *Electron. Lett.* **29**, 1807–1808 (1993)
- 30.13 S. P. Beeby, N. M. White: *Sensors Actuat. A* **88**, 189–197 (2001)
- 30.14 J. N. Ross: *Meas. Sci. Technol.* **6**, 405–409 (1995)
- 30.15 M. Prudenziati, B. Morten: *Microelectron. J.* **23**, 133–141 (1992)
- 30.16 G. Martinelli, M. C. Carotta: *Sensors Actuat. B* **23**, 157–161 (1995)
- 30.17 B. Morten, M. Prudenziati, F. Sirotti, G. De Cicco, A. Alberigi-Quaranta, L. Olumekor: *J. Mater. Sci. Mater. El.* **1**, 118–122 (1990)
- 30.18 N. Nakayama, H. Matsumoto, A. Nakano, S. Ikegami, H. Uda, T. Yamashita: *Jpn. J. Appl. Phys.* **19**, 703–712 (1980)

Part D Materials

Part D Materials for Optoelectronics and Photonics

31 III–V Ternary and Quaternary Compounds

Sadao Adachi, Gunma, Japan

32 Group III Nitrides

Ali Teke, Balikesir, Turkey
Hadis Morkoç, Richmond, USA

33 Electron Transport Within the III–V Nitride Semiconductors, GaN, AlN, and InN: A Monte Carlo Analysis

Brian E. Foutz, Endicott, USA
Stephen K. O'Leary, Regina, Canada
Michael Shur, Troy, USA
Lester F. Eastman, Ithaca, USA

34 II–IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe

Jifeng Wang, Sendai, Japan
Minoru Isshiki, Sendai, Japan

35 Doping Aspects of Zn–Based Wide-Band-Gap Semiconductors

Gertrude F. Neumark, New York, USA
Yinyan Gong, New York, USA
Igor L. Kuskovsky, Flushing, USA

36 II–VI Narrow-Bandgap Semiconductors for Optoelectronics

Ian M. Baker, Southampton, UK

37 Optoelectronic Devices and Materials

Stephen Sweeney, Guildford, UK
Alfred Adams, Surrey, UK

38 Liquid Crystals

David Dunmur, Southampton, UK
Geoffrey Luckhurst, Southampton, UK

39 Organic Photoconductors

David S. Weiss, Rochester, USA
Martin Abkowitz, Webster, USA

40 Luminescent Materials

Andy Edgar, Wellington, New Zealand

41 Nano-Engineered Tunable Photonic Crystals in the Near-IR and Visible Electromagnetic Spectrum

Harry Ruda, Toronto, Canada
Naomi Matsuura, Toronto, Canada

42 Quantum Wells, Superlattices, and Band-Gap Engineering

Mark Fox, Sheffield, UK

43 Glasses for Photonic Integration

Ray DeCorby, Edmonton, Canada

44 Optical Nonlinearity in Photonic Glasses

Keiji Tanaka, Sapporo, Japan

45 Nonlinear Optoelectronic Materials

Lukasz Brzozowski, Toronto, ON, Canada
Edward Sargent, Toronto, Canada

31. III–V Ternary and Quaternary Compounds

III–V ternary and quaternary alloy systems are potentially of great importance for many high-speed electronic and optoelectronic devices, because they provide a natural means of tuning the magnitude of forbidden gaps so as to optimize and widen the applications of such semiconductor devices. Literature on the fundamental properties of these material systems is growing rapidly. Even though the basic semiconductor alloy concepts are understood at this time, some practical and device parameters in these material systems have been hampered by a lack of definite knowledge of many material parameters and properties.

This chapter attempts to summarize, in graphical and tabular forms, most of the important theoretical and experimental data on the III–V ternary and quaternary alloy parameters and properties. They can be classified into six groups: (1) Structural parameters; (2) Mechanical, elastic, and lattice vibronic properties; (3) Thermal properties; (4) Energy band parameters; (5) Optical properties, and; (6) Carrier transport properties. The III–V ternary and quaternary alloys considered here are those of Group III (Al, Ga, In) and V (N, P, As, Sb) atoms. The model used in some cases is based on an interpolation scheme and, therefore, requires that data on the material parameters for the related binaries (AlN, AlP, GaN, GaP, etc.) are known. These data have been taken mainly from the Landolt–Börnstein collection, Vol. III/41, and from the *Handbook on Physical Properties of Semiconductors Volume 2: III–V Compound Semiconductors*, published by Springer in 2004. The material parameters and properties derived

31.1	Introduction to III–V Ternary and Quaternary Compounds	735
31.2	Interpolation Scheme	736
31.3	Structural Parameters	737
31.3.1	Lattice Parameters and Lattice–Matching Conditions Between III–V Quaternaries and Binary Substrates.....	737
31.3.2	Molecular and Crystal Densities...	737
31.4	Mechanical, Elastic and Lattice Vibronic Properties	739
31.4.1	Microhardness	739
31.4.2	Elastic Constants and Related Moduli	739
31.4.3	Long-Wavelength Phonons	739
31.5	Thermal Properties	741
31.5.1	Specific Heat and Debye Temperature	741
31.5.2	Thermal Expansion Coefficient	741
31.5.3	Thermal Conductivity	741
31.6	Energy Band Parameters	743
31.6.1	Bandgap Energy.....	743
31.6.2	Carrier Effective Mass	744
31.6.3	Deformation Potential	746
31.7	Optical Properties	748
31.7.1	The Reststrahlen Region.....	748
31.7.2	The Interband Transition Region .	749
31.8	Carrier Transport Properties	750
	References	751

here are used with wide success to obtain the general properties of these alloy semiconductors.

31.1 Introduction to III–V Ternary and Quaternary Compounds

III–V semiconducting compound alloys are widely used as materials for optoelectronic devices such as light-emitting diodes, laser diodes and photodetectors, as well as for electronic transport devices such as field effect transistors, high electron mobility transistors and het-

erojunction bipolar transistors. In a ternary alloy, the bandgap energy E_g and the lattice parameter a are generally both functions of a single composition parameter, so they cannot be selected independently. In quaternary alloys, on the other hand, the two com-

position parameters allow E_g and a to be selected independently, within the constraints of a given alloy–substrate system. Even though the basic semiconductor alloy concepts are understood at this time, the determination of some practical device parameters has been hampered by a lack of definite knowledge of many material parameters. This chapter provides data on the fundamental material properties of III–V ternary and quaternary alloys. The model used here is based on an interpolation scheme and thus requires that

values of the material parameters for the related end-point binaries are known. We therefore begin with the constituent binaries and gradually move on to alloys. The phenomenon of spontaneous ordering in semiconductor alloys, which can be categorized as a self-organized process, is observed to occur spontaneously during the epitaxial growth of certain alloys, and results in modifications to their structural, electronic and optical properties. This topic is omitted from the coverage [31.1].

31.2 Interpolation Scheme

The electronic energy band parameters of III–V compound alloys and their dependence on alloy composition are very important device parameters, and so they have received considerable attention in the past. Investigations of many device parameters have, however, been hampered by a lack of definite knowledge of various material parameters. This necessitates the use of some kind of interpolation scheme. Although the interpolation scheme is still open to experimental verification, it can provide more useful and reliable material parameters over the entire range of alloy composition [31.2].

If one uses the linear interpolation scheme, the ternary parameter T can be derived from the binary parameters (B) by

$$T_{A_xB_{1-x}C} = xB_{AC} + (1-x)B_{BC} \equiv a + bx \quad (31.1)$$

for an alloy of the form $A_xB_{1-x}C$, where $a \equiv B_{BC}$ and $b \equiv B_{AC} - B_{BC}$. Some material parameters, however, deviate significantly from the linear relation (31.1), and exhibit an approximately quadratic dependence on the mole fraction x . The ternary material parameter in such a case can be very efficiently approximated by the relationship

$$T_{A_xB_{1-x}C} = xB_{AC} + (1-x)B_{BC} + C_{A-B}x(1-x) \equiv a + bx + cx^2, \quad (31.2)$$

where $a \equiv B_{BC}$ and $b \equiv B_{AC} - B_{BC} + C_{A-B}$, and $c \equiv -C_{A-B}$. The parameter c is called the bowing or nonlinear parameter.

The quaternary material $A_xB_{1-x}C_yD_{1-y}$ is thought to be constructed from four binaries: AC, AD, BC, and BD. If one uses the linear interpolation scheme, the quaternary parameter Q can be derived from the B s by

$$Q(x, y) = xyB_{AC} + x(1-y)B_{AD} + (1-x)yB_{BC} + (1-x)(1-y)B_{BD}. \quad (31.3)$$

If one of the four binary parameters (e.g., B_{AD}) is lacking, Q can be estimated from

$$Q(x, y) = xB_{AC} + (y-x)B_{BC} + (1-y)B_{BD}. \quad (31.4)$$

The quaternary material $A_xB_yC_{1-x-y}D$ is thought to be constructed from three binaries: AD, BD, and CD. The corresponding linear interpolation is given by

$$Q(x, y) = xB_{AD} + yB_{BD} + (1-x-y)B_{CD}. \quad (31.5)$$

If the material parameter can be given by a specific expression owing to some physical basis, it is natural to consider that the interpolation scheme may also obey this expression. The static dielectric constant ϵ_s is just the case that follows the Clausius–Mosotti relation. Then, the interpolation expression for the $A_xB_{1-x}C_yD_{1-y}$ quaternary, for example, has the form

$$\begin{aligned} \frac{\epsilon_s(x, y) - 1}{\epsilon_s(x, y) - 2} = & xy \frac{\epsilon_s(AC) - 1}{\epsilon_s(AC) - 2} + x(1-y) \frac{\epsilon_s(AD) - 1}{\epsilon_s(AD) - 2} \\ & + (1-x)y \frac{\epsilon_s(BC) - 1}{\epsilon_s(BC) - 2} \\ & + (1-x)(1-y) \frac{\epsilon_s(BD) - 1}{\epsilon_s(BD) - 2}. \end{aligned} \quad (31.6)$$

When bowing from the anion sublattice disorder is independent of the disorder in the cation sublattice, the interpolation scheme is written by incorporating these cation and anion bowing parameters into the linear interpolation scheme as

$$\begin{aligned} Q(x, y) = & xyB_{AC} + x(1-y)B_{AD} + (1-x)yB_{BC} \\ & + (1-x)(1-y)B_{BD} + C_{A-B}x(1-x) \\ & + C_{C-D}y(1-y) \end{aligned} \quad (31.7)$$

for the $A_xB_{1-x}C_yD_{1-y}$ quaternary, or

$$\begin{aligned} Q(x, y) = & xB_{AD} + yB_{BD} + (1-x-y)B_{CD} \\ & + C_{A-B} - C_{xy}(1-x-y) \end{aligned} \quad (31.8)$$

for the $A_xB_yC_{1-x-y}D$ quaternary.

If relationships for the ternary parameters T s are available, the quaternary parameter Q can be expressed either as $(A_xB_{1-x}C_yD_{1-y})$

$$Q(x, y) = \frac{x(1-x)[yT_{ABC}(x) + (1-y)T_{ABD}(x)]}{x(1-x) + y(1-y)} + \frac{y(1-y)[xT_{ACD}(y) + (1-x)T_{BCD}(y)]}{x(1-x) + y(1-y)}, \quad (31.9)$$

or $(A_xB_yC_{1-x-y}D)$

$$Q(x, y) = \frac{xyT_{ABD}(u) + y(1-x-y)T_{BCD}(v)}{xy + y(1-x-y) + x(1-x-y)} + \frac{x(1-x-y)T_{ACD}(w)}{xy + y(1-x-y) + x(1-x-y)} \quad (31.10)$$

with

$$u = (1-x-y)/2, \quad v = (2-x-2y)/2, \quad w = (2-2x-y)/2. \quad (31.11)$$

31.3 Structural Parameters

31.3.1 Lattice Parameters and Lattice-Matching Conditions Between III–V Quaternaries and Binary Substrates

The lattice parameter a (c) is known to obey Vegard's law well, i. e., to vary linearly with composition. Thus, the lattice parameter for a III–V ternary can be simply obtained from (31.1) using the binary data

Table 31.1 Lattice parameters a and c and crystal density g for some III–V binaries at 300 K

Binary	Zinc blende a (Å)	Wurtzite a (Å)	c (Å)	g (g/cm ³)
AlN	–	3.112	4.982	3.258
AlP	5.4635	–	–	2.3604
AlAs	5.661 39	–	–	3.7302
AlSb	6.1355	–	–	4.2775
α -GaN	–	3.1896	5.1855	6.0865
β -GaN	4.52	–	–	6.02
GaP	5.4508	–	–	4.1299
GaAs	5.653 30	–	–	5.3175
GaSb	6.095 93	–	–	5.6146
InN	–	3.548	5.760	6.813
InP	5.8690	–	–	4.7902
InAs	6.0583	–	–	5.6678
InSb	6.479 37	–	–	5.7768

listed in Table 31.1 [31.3, 4]. Introducing the lattice parameters in Table 31.1 into (31.3) [(31.5)], one can also obtain the lattice-matching conditions for $A_{1-x}B_xC_yD_{1-y}$ ($A_xB_yC_{1-x-y}D$) quaternaries on various III–V binary substrates (GaAs, GaSb, InP and InAs). These results are summarized in Tables 31.2, 31.3, 31.4 and 31.5.

31.3.2 Molecular and Crystal Densities

The molecular density d_M can be obtained via

$$d_M = \frac{4}{a^3} \quad (31.12)$$

for zinc blende-type materials, and

$$d_M = \frac{4}{a_{\text{eff}}^3} \quad (31.13)$$

for wurtzite-type materials, where a_{eff} is an effective cubic lattice parameter defined by

$$a_{\text{eff}} = (\sqrt{3}a^2c)^{1/3}. \quad (31.14)$$

The X-ray crystal density g can be simply written, using d_M , as

$$g = \frac{Md_M}{N_A}, \quad (31.15)$$

Table 31.2 Lattice-matching conditions for some III–V quaternaries of type $A_xB_{1-x}C_yD_{1-y}$ at 300 K. $x = \frac{A_0+B_0y}{C_0+D_0y}$

Quaternary	Substrate	A_0	B_0	C_0	D_0	Remark
$Ga_xIn_{1-x}P_yAs_{1-y}$	GaAs	0.4050	–0.1893	0.4050	0.0132	$0 \leq y \leq 1.0$
	InP	0.1893	–0.1893	0.4050	0.0132	$0 \leq y \leq 1.0$
$Al_xIn_{1-x}P_yAs_{1-y}$	GaAs	0.4050	–0.1893	0.3969	0.0086	$0.04 \leq y \leq 1.0$
	InP	0.1893	–0.1893	0.3969	0.0086	$0 \leq y \leq 1.0$

where M is the molecular weight and $N_A = 6.022 \times 10^{23} \text{ mole}^{-1}$ is the Avogadro constant. We list d_M and g for some III–V binaries in Table 31.1. Alloy values of d_M and g can be accurately obtained using Vegard’s law, i.e., (31.1), (31.3), and (31.5).

Table 31.3 Lattice-matching conditions for some III–V quaternaries of type $A_xB_{1-x}C_yD_{1-y}$ at 300 K. $y = \frac{A_0+B_0x}{C_0+D_0x}$

Quaternary	Substrate	A_0	B_0	C_0	D_0	Remark
$Al_xGa_{1-x}P_yAs_{1-y}$	GaAs	0	0.0081	0.2025	−0.0046	$0 \leq x \leq 1.0$
$Al_xGa_{1-x}As_ySb_{1-y}$	GaSb	0	0.0396	0.4426	0.0315	$0 \leq x \leq 1.0$
	InP	0.2269	0.0396	0.4426	0.0315	$0 \leq x \leq 1.0$
	InAs	0.0376	0.0396	0.4426	0.0315	$0 \leq x \leq 1.0$
$Al_xGa_{1-x}P_ySb_{1-y}$	GaAs	0.4426	0.0396	0.6451	0.0269	$0 \leq x \leq 1.0$
	GaSb	0	0.0396	0.6451	0.0269	$0 \leq x \leq 1.0$
	InP	0.2269	0.0396	0.6451	0.0269	$0 \leq x \leq 1.0$
$Ga_xIn_{1-x}As_ySb_{1-y}$	InAs	0.0376	0.0396	0.6451	0.0269	$0 \leq x \leq 1.0$
	GaSb	0.3834	−0.3834	0.4211	0.0216	$0 \leq x \leq 1.0$
	InP	0.6104	−0.3834	0.4211	0.0216	$0.47 \leq x \leq 1.0$
$Ga_xIn_{1-x}P_ySb_{1-y}$	InAs	0.4211	−0.3834	0.4211	0.0216	$0 \leq x \leq 1.0$
	GaAs	0.8261	−0.3834	0.6104	0.0348	$0.52 \leq x \leq 1.0$
	GaSb	0.3834	−0.3834	0.6104	0.0348	$0 \leq x \leq 1.0$
$Al_xIn_{1-x}As_ySb_{1-y}$	InP	0.6104	−0.3834	0.6104	0.0348	$0 \leq x \leq 1.0$
	InAs	0.4211	−0.3834	0.6104	0.0348	$0 \leq x \leq 1.0$
	GaSb	0.3834	−0.3439	0.4211	0.0530	$0 \leq x \leq 1.0$
$Al_xIn_{1-x}P_ySb_{1-y}$	InP	0.6104	−0.3439	0.4211	0.0530	$0.48 \leq x \leq 1.0$
	InAs	0.4211	−0.3439	0.4211	0.0530	$0 \leq x \leq 1.0$
	GaAs	0.8261	−0.3439	0.6104	0.0616	$0.53 \leq x \leq 1.0$
	GaSb	0.3834	−0.3439	0.6104	0.0616	$0 \leq x \leq 1.0$
	InP	0.6104	−0.3439	0.6104	0.0616	$0 \leq x \leq 1.0$
	InAs	0.4211	−0.3439	0.6104	0.0616	$0 \leq x \leq 1.0$

Table 31.4 Lattice-matching conditions for some III–V quaternaries of type $A_xB_yC_{1-x-y}D$ at 300 K. $y = A_0 + B_0x$

Quaternary	Substrate	A_0	B_0	Remark
$Al_xGa_yIn_{1-x-y}P$	GaAs	0.5158	−0.9696	$0 \leq x \leq 0.53$
$Al_xGa_yIn_{1-x-y}As$	InP	0.4674	−0.9800	$0 \leq x \leq 0.48$

Table 31.5 Lattice-matching conditions for some III–V quaternaries of type $AB_xC_yD_{1-x-y}$ at 300 K. $x = A_0 + B_0y$

Quaternary	Substrate	A_0	B_0	Remark
$AlP_xAs_ySb_{1-x-y}$	GaAs	0.7176	−0.7055	$0 \leq y \leq 0.96$
	InP	0.3966	−0.7055	$0 \leq y \leq 0.56$
	InAs	0.1149	−0.7055	$0 \leq y \leq 0.16$
$GaP_xAs_ySb_{1-x-y}$	GaAs	0.6861	−0.6861	$0 \leq y \leq 1.0$
	InP	0.3518	−0.6861	$0 \leq y \leq 0.51$
	InAs	0.0583	−0.6861	$0 \leq y \leq 0.085$
$InP_xAs_ySb_{1-x-y}$	GaSb	0.6282	−0.6899	$0 \leq y \leq 0.911$
	InAs	0.6899	−0.6899	$0 \leq y \leq 1.0$

31.4 Mechanical, Elastic and Lattice Vibronic Properties

31.4.1 Microhardness

The hardness test has been used for a long time as a simple means of characterizing the mechanical behavior of solids. The Knoop hardness H_P for $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$ lattice-matched to InP has been reported [31.5], and is found to increase gradually from 520 kg/mm² for $y = 0$ ($\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$) to 380 kg/mm² for $y = 1.0$ (InP). It has also been reported that the microhardness in $\text{Al}_x\text{Ga}_{1-x}\text{N}$ thin film slightly decreases with increasing AlN composition x [31.6].

31.4.2 Elastic Constants and Related Moduli

Although the elastic properties of the III–V binaries have been studied extensively, little is known about their alloys. Recent studies, however, suggested that the elastic properties of the alloys can be obtained, to a good approximation, by averaging the binary endpoint values [31.7, 8]. We have, therefore, listed in Tables 31.6 and 31.7 the elastic stiffness (C_{ij}) and compliance constants (S_{ij}) for some III–V binaries with zinc blende and wurtzite structures, respectively. Table 31.8 also sum-

marizes the functional expressions for the bulk modulus B_u , Young's modulus Y , and Poisson's ratio P . Note that Y and P are not isotropic, even in the cubic zinc blende lattice.

31.4.3 Long-Wavelength Phonons

The atoms of a crystal can be visualized as being joined by harmonic springs, and the crystal dynamics can be analyzed in terms of a linear combination of $3N$ normal modes of vibration (N is the number of different types of atoms; different in terms of mass or ordering in space). In alloys, the nature of the lattice optical spectrum depends on the difference between the quantities representing the lattice vibronic properties of the components. If these quantities are similar, then the optical response of an alloy is similar to the response of a crystal with the quantities averaged over the composition (one-mode behavior). In one-mode systems, such as most I–VII alloys, a single set of long-wavelength optical modes appears, as schematically shown in Fig. 31.1. When the parameters differ strongly, the response of a system is more complex; the spectrum contains a num-

Table 31.6 Elastic stiffness (C_{ij}) and compliance constants (S_{ij}) for some cubic III–V binaries at 300 K

Binary	C_{ij} (10^{11} dyn/cm ²)			S_{ij} (10^{-12} cm ² /dyn)		
	C_{11}	C_{12}	C_{44}	S_{11}	S_{12}	S_{44}
AlP	15.0*	6.42*	6.11*	0.897*	−0.269*	1.64*
AlAs	11.93	5.72	5.72	1.216	−0.394	1.748
AlSb	8.769	4.341	4.076	1.697	−0.5618	2.453
β -GaN	29.1*	14.8*	15.8*	0.523*	−0.176*	0.633*
GaP	14.050	6.203	7.033	0.9756	−0.2988	1.422
GaAs	11.88	5.38	5.94	1.173	−0.366	1.684
GaSb	8.838	4.027	4.320	1.583	−0.4955	2.315
InP	10.22	5.73	4.42	1.639	−0.589	2.26
InAs	8.329	4.526	3.959	1.945	−0.6847	2.526
InSb	6.608	3.531	3.027	2.410	−0.8395	3.304

* Theoretical

Table 31.7 Elastic stiffness (C_{ij}) and compliance constants (S_{ij}) for some wurtzite III–V binaries at 300 K

Binary	C_{ij} (10^{11} dyn/cm ²)						S_{ij} (10^{-12} cm ² /dyn)					
	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	C_{66}^*	S_{11}	S_{12}	S_{13}	S_{33}	S_{44}	S_{66}^{*2}
AlN	41.0	14.0	10.0	39.0	12.0	13.5	0.285	−0.085	−0.051	0.283	0.833	0.740
α -GaN	37.3	14.1	8.0	38.7	9.4	11.6	0.320	−0.112	−0.043	0.276	1.06	0.864
InN	19.0	10.4	12.1	18.2	0.99	4.3	0.957	−0.206	−0.499	1.21	10.1	2.33

*¹ $C_{66} = 1/2(C_{11} - C_{12})$, *² $S_{66} = 2(S_{11} - S_{12})$

Table 31.8 Functional expressions for the bulk modulus B_u , Young’s modulus Y , and Poisson’s ratio P in semiconductors with zinc blende (ZB) and wurtzite (W) structures

Parameter	Structure	Expression	Remark
B_u	ZB	$(C_{11} + 2C_{12})/3$	
	W	$[(C_{11} + C_{12})C_{33} - 2C_{13}^2]/(C_{11} + C_{12} + 2C_{33} - 4C_{13})$	
Y	ZB	$1/S_{11}$	(100), [001]
		$1/(S_{11} - S/2)$	(100), [011]
		$1/S_{11}$	(110), [001]
		$1/(S_{11} - 2S/3)$	(110), [111]
		$1/(S_{11} - S/2)$	(111)
	W	$1/S_{11}$	$c \perp \boldsymbol{l}$
		$1/S_{33}$	$c \parallel \boldsymbol{l}$
P	ZB	$-S_{12}/S_{11}$	(100), $m = [010]$, $n = [001]$
		$-(S_{12} + S/2)/(S_{11} - S/2)$	(100), $m = [011]$, $n = [0\bar{1}1]$
		$-S_{12}/S_{11}$	(110), $m = [001]$, $n = [1\bar{1}0]$
		$-(S_{12} + S/3)/(S_{11} - 2S/3)$	(110), $m = [1\bar{1}1]$, $n = [1\bar{1}2]$
		$-(S_{12} + S/6)/(S_{11} - S/2)$	(111)
	W	$(1/2)[1 - (Y/3B_u)]$	$c \perp \boldsymbol{l}$, $c \parallel \boldsymbol{l}$

$S = S_{11} - S_{12} - (S_{44}/2)$; \boldsymbol{l} = directional vector; m = direction for a longitudinal stress; n = direction for a transverse strain ($n \perp m$)

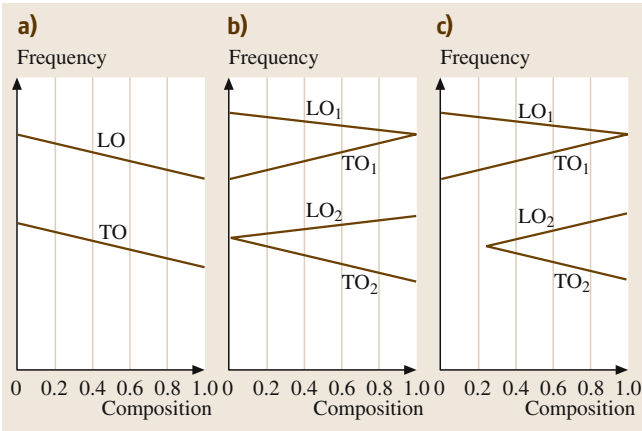


Fig. 31.1a–c Three different types of long-wavelength phonon mode behavior in ternary alloys: (a) one-mode; (b) two-mode; and (c) one-two-mode ◀

cal modes with frequencies characteristic of each end member and strengths that are roughly proportional to the respective concentrations.

As seen in Table 31.9, the long-wavelength optical phonons in III–V ternaries exhibit either one-mode or two-mode behavior, or more rigorously, three different types of mode behavior: one-mode, two-mode, and one-two-mode behaviors. The one-two-mode system exhibits a single mode over only a part of the composition range, with two modes observed over the remaining range of compositions.

In a quaternary alloy of the $A_xB_{1-x}C_yD_{1-y}$ type, there are four kinds of unit cells: AC, AD, BC, and BD. On the other hand, in the $A_xB_yC_{1-x-y}D$ type there are three kinds of unit cells: AD, BD, and CD. We can, thus, expect four-mode or three-

Table 31.9 Behavior of the long-wavelength optical modes in III–V ternary and quaternary alloys

Behavior	Alloy
One mode	AlGa N (LO), AlIn N , GaIn N , AlAs Sb
Two mode	AlGa N (TO), AlGa P , AlGa As , AlGa Sb , AlIn As , AlIn Sb , GaIn P , GaIn As , Ga NAs , Ga PAs , Ga PSb
One–two mode	AlIn P , GaIn Sb , InAs Sb
Three mode	AlGa $AsSb$, GaIn $AsSb$, AlGaIn P , AlGaIn As , In $PAsSb$
Four mode	GaIn PSb , GaIn PAs

mode behavior of the long-wavelength optical modes in such quaternary alloys ([31.9]; Table 31.9). However, the $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{Sb}_{1-y}$ quaternary showed three-mode behavior with GaAs, InSb and mixed InAs/GaAs characteristics [31.10]. The $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{Sb}_{1-y}$ quaternary was also reported to show two-mode or three-mode behavior, depending on the alloy composition [31.11].

The long-wavelength optical phonon behavior in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary has been studied both theoretically and experimentally. These studies suggest that the optical phonons in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ exhibit the two-mode behavior over the whole composition range. Thus, the

$\text{Al}_x\text{Ga}_{1-x}\text{As}$ system has two couples of the transverse optical (TO) and longitudinal optical (LO) modes; one is the GaAs-like mode and the other is the AlAs-like mode. Each phonon frequency can be expressed as [31.12]

- TO (GaAs): $268 - 14x \text{ cm}^{-1}$,
- LO (GaAs): $292 - 38x \text{ cm}^{-1}$,
- TO (AlAs): $358 + 4x \text{ cm}^{-1}$,
- LO (AlAs): $358 + 71x - 26x^2 \text{ cm}^{-1}$.

It is observed that only the AlAs-like LO mode shows a weak nonlinearity with respect to the alloy composition x .

31.5 Thermal Properties

31.5.1 Specific Heat and Debye Temperature

Since alloying has no significant effect on elastic properties, it appears that using the linear interpolation scheme for alloys can provide generally acceptable specific heat values (C). In fact, it has been reported that the C values for $\text{InP}_x\text{As}_{1-x}$ [31.13] and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ [31.14] vary fairly linearly with alloy composition x . It has also been shown [31.12] that the Debye temperature θ_D for alloys shows very weak nonlinearity with composition. From these facts, one can suppose that the linear interpolation scheme may provide generally acceptable C and θ_D values for III-V semiconductor alloys. We have, therefore, listed in Table 31.10 the III-V binary endpoint

values for C and θ_D at $T = 300 \text{ K}$. Using these values, the linearly interpolated C value for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ can be obtained from $C(x) = 0.424x + 0.327(1-x) = 0.327 + 0.097x \text{ (J/gK)}$.

31.5.2 Thermal Expansion Coefficient

The linear thermal expansion coefficient α_{th} is usually measured by measuring the temperature dependence of the lattice parameter. The composition dependence of α_{th} has been measured for many semiconductor alloys, including $\text{Ga}_x\text{In}_{1-x}\text{P}$ [31.15] and $\text{GaP}_x\text{As}_{1-x}$ [31.16]. These studies indicate that the α_{th} value varies almost linearly with composition. This suggests that the thermal expansion coefficient can be accurately estimated using linear interpolation. In fact, we plot in Fig. 31.2 the 300 K value of α_{th} as a function of x for the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary. By using the least-squares fit procedure, we obtain the linear relationship between α_{th} and x as $\alpha_{\text{th}}(x) = 6.01 - 1.74x \text{ (} 10^{-6} \text{ K}^{-1} \text{)}$. This expression is almost the same as that obtained using the linear interpolation expression: $\alpha_{\text{th}}(x) = 4.28x + 6.03(1-x) = 6.03 - 1.75x \text{ (} 10^{-6} \text{ K}^{-1} \text{)}$. The binary endpoint values of α_{th} are listed in Table 31.10.

31.5.3 Thermal Conductivity

The lattice thermal conductivity κ , or the thermal resistivity $W = 1/\kappa$, results mainly from interactions between phonons and from the scattering of phonons by crystalline imperfections. It is important to point out that when large numbers of foreign atoms are added to the host lattice, as in alloying, the thermal conductivity may

Table 31.10 Specific heat C and Debye temperature θ_D for some III-V binaries at 300 K

Binary	$C \text{ (J/gK)}$	$\theta_D \text{ (K)}$	$\alpha_{\text{th}} \text{ (} 10^{-6} \text{ K}^{-1} \text{)}$
AlN	0.728	988	3.042 ($\perp c$), 2.227 ($\parallel c$)
AlP	0.727	687	
AlAs	0.424	450	4.28
AlSb	0.326* ¹	370* ¹	4.2
α -GaN	0.42	821	5.0 ($\perp c$), 4.5 ($\parallel c$)
GaP	0.313	493* ²	4.89
GaAs	0.327	370	6.03
GaSb	0.344* ¹	240* ¹	6.35
InN	2.274	674	3.830 ($\perp c$), 2.751 ($\parallel c$)
InP	0.322	420* ¹	4.56
InAs	0.352	280* ¹	≈ 5.0
InSb	0.350* ¹	161* ¹	5.04

*¹ At 273 K, *² at 150 K

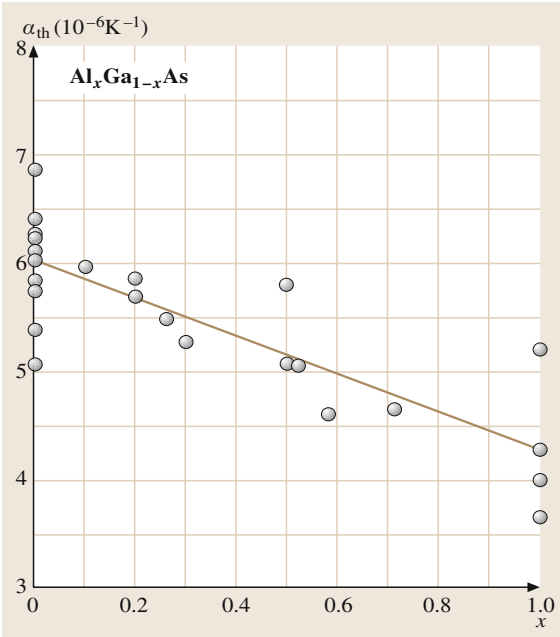


Fig. 31.2 Thermal expansion coefficient α_{th} as a function of x for the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary at $T = 300$ K. The experimental data are gathered from various sources. The solid line is linearly interpolated between the AlAs and GaAs values

decrease significantly. Experimental data on various alloy semiconductors, in fact, exhibit strong nonlinearity

Table 31.11 Thermal resistivity values W for some III–V binaries at 300 K. Several cation and anion bowing parameters used for the calculation of alloy values are also listed in the last column

Binary	W (cmK/W)	C_{A-B} (cmK/W)
AlN	0.31 ^{*1}	
AlP	1.11	
AlAs	1.10	
AlSb	1.75	
α -GaN	0.51 ^{*1}	$C_{\text{Al-Ga}} = 32$
GaP	1.30	$C_{\text{Ga-In}} = 72$
GaAs	2.22	$C_{\text{P-As}} = 25$
GaSb	2.78	$C_{\text{As-Sb}} = 90$
InN	2.22 ^{*2}	
InP	1.47	
InAs	3.33	
InSb	5.41–6.06	

^{*1} Heat flow parallel to the basal plane, ^{*2} ceramics

with respect to the alloy composition. Such a composition dependence can be successfully explained by using the quadratic expression of (31.2) or (31.6) [31.17].

In Fig. 31.3 we compare the results calculated from (31.2) [(31.7)] to the experimental data for $\text{Al}_x\text{Ga}_{1-x}\text{As}$, $\text{Al}_x\text{Ga}_{1-x}\text{N}$ and $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}/\text{InP}$ alloys. The binary W values used in these calculations are taken from Table 31.11. The corresponding nonlin-

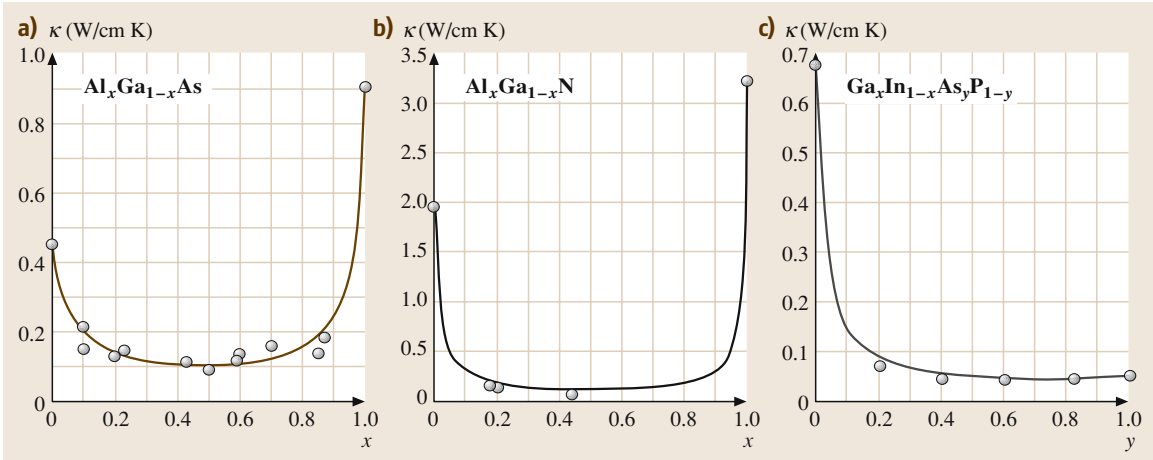


Fig. 31.3a–c Thermal conductivity κ as a function of $x(y)$ for (a) $\text{Al}_x\text{Ga}_{1-x}\text{As}$, (b) $\text{Al}_x\text{Ga}_{1-x}\text{N}$, and (c) $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$ lattice-matched to InP at $T = 300$ K. The experimental data (solid circles) are gathered from various sources. The solid lines represent the results calculated from (31.2) and (31.6) using the binary endpoint values and nonlinear parameters in Table 31.11

ear parameters C_{A-B} are also listed in Table 31.11. The agreement between the calculated and experimental data is excellent. By applying the present model, it is possi-

ble to estimate the κ (or W) values of experimentally unknown III–V alloy systems, such as $\text{GaAs}_x\text{Sb}_{1-x}$ and $\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{As}$.

31.6 Energy Band Parameters

31.6.1 Bandgap Energy

Lowest Direct and Lowest Indirect Band Gaps

The bandgap energies of III–V ternaries usually deviate from the simple linear relation of (31.1) and have an approximately quadratic dependence on the alloy composition x . Table 31.12 summarizes the lowest direct gap energy E_0 and the lowest indirect gap energies E_g^X and E_g^L for some III–V binaries of interest here. The corresponding nonlinear parameters C_{A-B} are listed in Table 31.13 [31.18]. Note that the E_g^X and E_g^L transitions correspond to those from the highest valence band at the Γ point to the lowest conduction band near X ($\Gamma_8 \rightarrow X_6$) or near L ($\Gamma_8 \rightarrow L_6$), respectively. The E_0 transitions take place at the Γ point ($\Gamma_8 \rightarrow \Gamma_6$).

Figure 31.4 plots the values of E_0 and E_g^X as a function of alloy composition x for the $\text{Ga}_x\text{In}_{1-x}\text{P}$ ternary at $T = 300\text{ K}$. The solid lines are obtained by introducing the numerical values from Tables 31.12 and 31.13 into (31.2). These curves provide the direct-indirect crossover composition at $x \approx 0.7$. Figure 31.5 also shows the variation in composition of E_0 in the

Table 31.12 Band-gap energies, E_0 , E_g^X and E_g^L , for some III–V binaries at 300 K. ZB = zinc blende

Binary	E_0 (eV)	E_g^X (eV)	E_g^L (eV)
AlN	6.2	–	–
AlN (ZB)	5.1	5.34	9.8*
AlP	3.91	2.48	3.30
AlAs	3.01	2.15	2.37
AlSb	2.27	1.615	2.211
α -GaN	3.420	–	–
β -GaN	3.231	4.2*	5.5*
GaP	2.76	2.261	2.63
GaAs	1.43	1.91	1.72
GaSb	0.72	1.05	0.76
InN	0.7–1.1	–	–
InP	1.35	2.21	2.05
InAs	0.359	1.37	1.07
InSb	0.17	1.63	0.93

* Theoretical

$\text{Ga}_x\text{In}_{1-x}\text{As}$, $\text{InAs}_x\text{Sb}_{1-x}$ and $\text{Ga}_x\text{In}_{1-x}\text{Sb}$ ternaries. It is understood from Table 31.13 that the bowing parameters for the bandgap energies of III–V ternaries are negative or very small, implying a downward bowing or a linear interpolation to within experimen-

Table 31.13 Bowing parameters used in the calculation of E_0 , E_g^X and E_g^L for some III–V ternaries. * W = wurtzite; ZB = zinc blende

Ternary	Bowing parameter C_{A-B} (eV)		
	E_0	E_g^X	E_g^L
(Al,Ga)N (W)	–1.0	–	–
(Al,Ga)N (ZB)	0	–0.61	–0.80
(Al,In)N (W)	$-16 + 9.1x$	–	–
(Al,In)N (ZB)	$-16 + 9.1x$		
(Ga,In)N (W)	–3.0	–	–
(Ga,In)N (ZB)	–3.0	–0.38	
(Al,Ga)P	0	–0.13	
(Al,In)P	–0.24	–0.38	
(Ga,In)P	–0.65	–0.18	–0.43
(Al,Ga)As	–0.37	–0.245	–0.055
(Al,In)As	–0.70	0	
(Ga,In)As	–0.477	–1.4	–0.33
(Al,Ga)Sb	–0.47	0	–0.55
(Al,In)Sb	–0.43		
(Ga,In)Sb	–0.415	–0.33	–0.4
Al(P,As)	–0.22	–0.22	–0.22
Al(P,Sb)	–2.7	–2.7	–2.7
Al(As,Sb)	–0.8	–0.28	–0.28
Ga(N,P) (ZB)	–3.9		
Ga(N,As) (ZB)	$-120.4 + 100x$		
Ga(P,As)	–0.19	–0.24	–0.16
Ga(P,Sb)	–2.7	–2.7	–2.7
Ga(As,Sb)	–1.43	–1.2	–1.2
In(N,P) (ZB)	–15		
In(N,As) (ZB)	–4.22		
In(P,As)	–0.10	–0.27	–0.27
In(P,Sb)	–1.9	–1.9	–1.9
In(As,Sb)	–0.67	–0.6	–0.6

* In those case where no value is listed, linear variation should be assumed

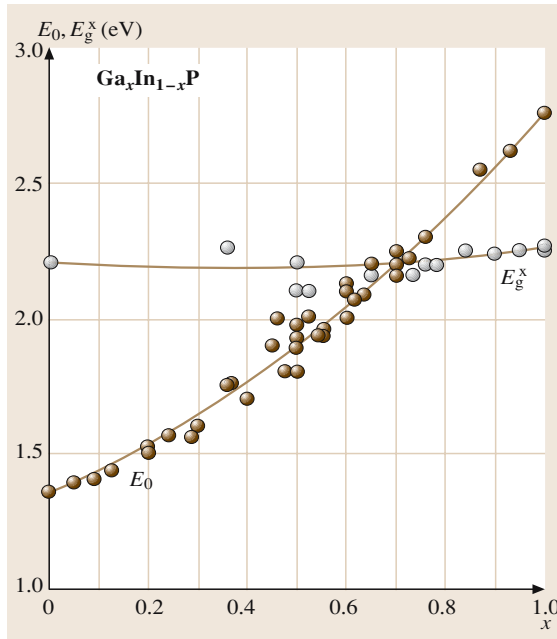


Fig. 31.4 Variation of the lowest direct gap (E_0) and lowest indirect gap energies (E_g^X) in the $\text{Ga}_x\text{In}_{1-x}\text{P}$ ternary at $T = 300$ K. The experimental data are gathered from various sources. The solid lines are calculated from (31.2) using the binary endpoint values and bowing parameters in Tables 31.12 and 31.13

Higher-Lying Band Gaps

The important optical transition energies observed at energies higher than E_0 are labeled E_1 and E_2 . We summarize in Table 31.15 the higher-lying bandgap energies E_1 and E_2 for some III–V binaries. The corresponding bowing parameters for these gaps are listed in Table 31.16.

31.6.2 Carrier Effective Mass

Electron Effective Mass

Since the carrier effective mass is strongly connected with the carrier mobility, it is known to be one of the most important device parameters. Effective masses can be measured by a variety of techniques, such as the Shubnikov-de Haas effect, magnetophonon resonance, cyclotron resonance, and interband magneto-optical effects. We list in Table 31.17 the electron effective mass (m_e^Γ) at the Γ -conduction band and the density of states (m_e^σ) and conductivity masses (m_e^σ) at the X-conduction and L-conduction bands of some III–V binaries. We also list in Table 31.18 the bowing parameters used when calculating the electron effective mass m_e^Γ for some III–V

tal uncertainty (Figs. 31.4, 31.5). It should be noted that nitrogen incorporation into (In,Ga)(P,As) results in a giant bandgap bowing of the host lattice for increasing nitrogen concentration [31.19]. We also summarize in Table 31.14 the expressions for the E_0 gap energy of some III–V quaternaries as a function of alloy composition.

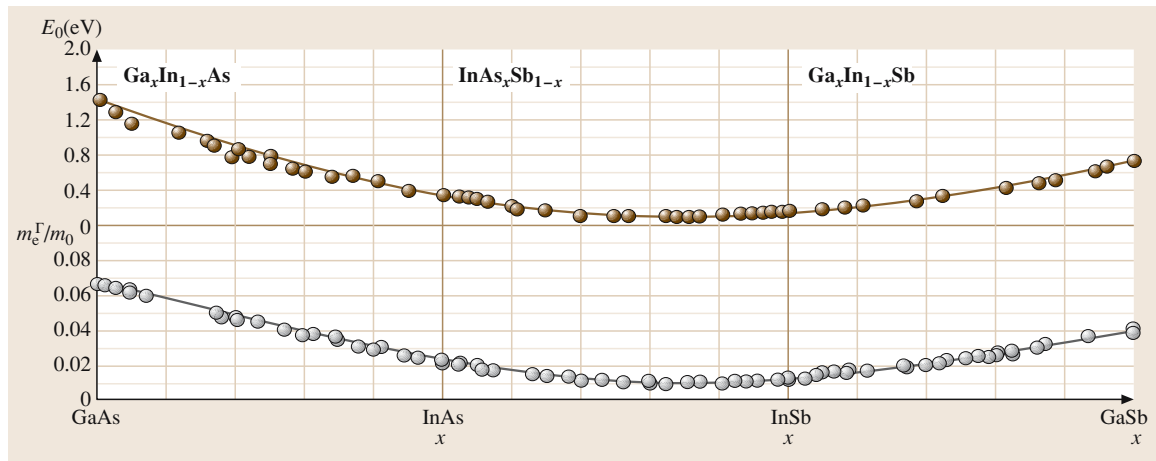


Fig. 31.5 Variation of the lowest direct gap energy E_0 ($T = 300$ K) and electron effective mass m_e^Γ at the Γ -conduction bands of $\text{Ga}_x\text{In}_{1-x}\text{As}$, $\text{InAs}_x\text{Sb}_{1-x}$ and $\text{Ga}_x\text{In}_{1-x}\text{Sb}$ ternaries. The experimental data are gathered from various sources. The solid lines are calculated from (31.2) using the binary endpoint values and bowing parameters in Tables 31.12 and 31.13 (E_0) and those in Tables 31.17 and 31.18 (m_e^Γ)

Table 31.14 Bandgap energies E_0 for some III–V quaternaries at 300 K

Quaternary	E_0 (eV)
$\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}/\text{InP}$	$0.75 + 0.48y + 0.12y^2$
$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{Sb}_{1-y}/\text{GaSb}$	$0.290 - 0.165x + 0.60x^2$
$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{Sb}_{1-y}/\text{InAs}$	$0.36 - 0.23x + 0.54x^2$
$\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{P}/\text{GaAs}^*$	$1.899 + 0.563x + 0.12x^2$
$\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{As}/\text{InP}$	$0.75 + 0.75x$
$\text{InP}_x\text{As}_y\text{Sb}_{1-x-y}/\text{InAs}$	$0.576 - 0.22y$

* The lowest indirect gap energy for this quaternary alloy can be obtained via $E_g^X = 2.20 - 0.09x$ eV

Table 31.15 Higher-lying bandgap energies, E_1 and E_2 , for some III–V binaries at 300 K

Binary	E_1 (eV)	E_2 (eV)
AlN	7.76	8.79
AlP	4.30	4.63
AlAs	3.62–3.90	4.853, 4.89
AlSb	2.78–2.890	4.20–4.25
α -GaN	6.9	8.0
β -GaN	7.0	7.6
GaP	3.71	5.28
GaAs	2.89–2.97	4.960–5.45
GaSb	2.05	4.08–4.20
InN	5.0	7.6
InP	3.17	4.70 (E'_0)
InAs	2.50	4.70
InSb	1.80	3.90

ternaries from (31.2). Note that the density of states mass m_e^α for electrons in the conduction band minima $\alpha = \Gamma$, X, and L can be obtained from

$$m_e^\alpha = N^{2/3} m_{t\alpha}^{2/3} m_{l\alpha}^{1/3}, \quad (31.16)$$

where N is the number of equivalent α minima ($N = 1$ for the Γ minimum, $N = 3$ for the X minima, and $N = 4$ for the L minima). The two masses m_l and m_t in (31.16) are called the longitudinal and transverse masses, respectively. The density of states effective mass m_e^α is used to calculate the density of states. The conductivity effective mass m_c^α , which can be used for calculating the conductivity (mobility), is also given by

$$m_c^\alpha = \frac{3m_{t\alpha}m_{l\alpha}}{m_{t\alpha} + 2m_{l\alpha}}. \quad (31.17)$$

Since $m_{t\Gamma} = m_{l\Gamma}$ at the $\alpha = \Gamma$ minimum of cubic semiconductors, we have the relation $m_c^\Gamma = m_l^\Gamma$. In the case of wurtzite semiconductors, we have the relation $m_c^\Gamma \neq m_l^\Gamma$, but the difference is very small.

Table 31.16 Bowing parameters used in the calculation of the higher-lying bandgap energies, E_1 and E_2 , for some cubic III–V ternaries

Ternary	C_{A-B} (eV)	
	E_1	E_2
(Al,Ga)P	0	0
(Al,In)P	0	0
(Ga,In)P	−0.86	0
(Al,Ga)As	−0.39	0
(Al,In)As	−0.38	
(Ga,In)As	−0.51	−0.27
(Al,Ga)Sb	−0.31	−0.34
(Al,In)Sb	−0.25	
(Ga,In)Sb	−0.33	−0.24
Ga(N,P)	0	0
Ga(N,As)	0	0
Ga(P,As)	0	0
Ga(As,Sb)	−0.59	−0.19
In(P,As)	−0.26	0
In(As,Sb)	≈ -0.55	≈ -0.6

* In those cases where no value is listed, linear variation should be assumed

The composition dependence of the electron effective mass m_e^Γ at the Γ -conduction bands of $\text{Ga}_x\text{In}_{1-x}\text{As}$, $\text{InAs}_x\text{Sb}_{1-x}$ and $\text{Ga}_x\text{In}_{1-x}\text{Sb}$ ternaries is plotted in Fig. 31.5. The solid lines are calculated from (31.2) using the binary endpoint values and bowing parameters in Tables 31.17 and 31.18. For conventional semiconductors, the values of the effective mass are known to decrease with decreasing bandgap energy (Fig. 31.5). This is in agreement with a trend predicted by the $\mathbf{k} \cdot \mathbf{p}$ theory [31.2]. In III–V–N alloys, the electron effective mass has been predicted to increase with increasing nitrogen composition in the low composition range [31.19]. This behavior is rather unusual, and in fact is opposite to what is seen in conventional semiconductors. However, a more recent study suggested that the effective electron mass in $\text{GaN}_x\text{As}_{1-x}$ decreases from $0.084m_0$ to $0.029m_0$ as x increases from 0 to 0.004 [31.20]. We also summarize in Table 31.19 the composition dependence of m_e^Γ , determined for $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$ and $\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{As}$ quaternaries lattice-matched to InP.

Hole Effective Mass

The effective mass can only be clearly defined for an isotropic parabolic band. In the case of III–V materials, the valence bands are warped from spherical symmetry some distance away from the Brillouin zone center

Table 31.17 Electron effective mass at the Γ -conduction band (m_e^Γ) and density of states (m_e^α) and conductivity masses (m_e^α) at the X-conduction and L-conduction bands of some III–V binaries. ZB = zinc blende

Binary	m_e^Γ/m_0	Density of states mass		Conductivity mass	
		m_e^X/m_0	m_e^L/m_0	m_e^X/m_0	m_e^L/m_0
AlN	0.29*	–	–	–	–
AlN (ZB)	0.26*	0.78*		0.37*	
AlP	0.220*	1.14*		0.31*	
AlAs	0.124	0.71	0.78	0.26*	0.21*
AlSb	0.14	0.84	1.05*	0.29	0.28*
α -GaN	0.21	–	–	–	–
β -GaN	0.15	0.78*		0.36*	
GaP	0.114	1.58	0.75*	0.37	0.21*
GaAs	0.067	0.85	0.56	0.32	0.11
GaSb	0.039	1.08*	0.54	0.44*	0.12
InN	0.07	–	–	–	–
InP	0.079 27	1.09*	0.76*	0.45*	0.19*
InAs	0.024	0.98*	0.94*	0.38*	0.18*
InSb	0.013				
* Theoretical					

Table 31.18 Bowing parameter used in the calculation of the electron effective mass m_e^Γ at the Γ -conduction bands of some III–V ternaries

Ternary	$C_{A-B}(m_0)$
(Ga,In)P	−0.019
(Al,In)As	0
(Ga,In)As	−0.0049
(Ga,In)Sb	−0.0092
Ga(P,As)	0
In(P,As)	0
In(As,Sb)	−0.030

Table 31.19 Electron effective mass m_e^Γ at the Γ -conduction bands of some III–V quaternaries

Quaternary	m_e^Γ/m_0
$\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}/\text{InP}$	$0.043 + 0.036y$
$\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{As}/\text{InP}$	$0.043 + 0.046x - 0.017x^2$

(Γ). Depending on the measurement or calculation technique employed, different values of hole masses are then possible experimentally or theoretically. Thus, it is always important to choose the correct definition of the effective hole mass which appropriate to the physical phenomenon considered.

We list in Table 31.20 the density of states heavy hole (m_{HH}^*), the averaged light hole (m_{LH}^*), and spin orbit splitoff effective hole masses (m_{SO}) in some cubic

III–V semiconductors. These masses are, respectively, defined using Luttinger’s valence band parameters γ_i by

$$m_{\text{HH}}^* = \frac{(1 + 0.05\gamma_h + 0.0164\gamma_h^2)^{2/3}}{\gamma_1 - \bar{\gamma}}, \quad (31.18)$$

$$m_{\text{LH}}^* = \frac{1}{\gamma_1 + \bar{\gamma}}, \quad (31.19)$$

$$m_{\text{SO}} = \frac{1}{\gamma_1} \quad (31.20)$$

with

$$\bar{\gamma} = (2\gamma_2^2 + 2\gamma_3^2)^{1/2}, \quad \gamma_h = \frac{6(\gamma_3^2 - \gamma_2^2)}{\bar{\gamma}(\gamma_1 - \bar{\gamma})}. \quad (31.21)$$

Only a few experimental studies have been performed on the effective hole masses in III–V alloys, e.g., the $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$ quaternary [31.2]. While some data imply a bowing parameter, the large uncertainties in existing determinations make it difficult to conclusively state that such experimental values are preferable to a linear interpolation. The binary endpoint data listed in Table 31.20 enable us to estimate alloy values using the linear interpolation scheme.

31.6.3 Deformation Potential

The deformation potentials of the electronic states at the Brillouin zone centers of semiconductors play an important role in many physical phenomena. For example,

Table 31.20 Density of states heavy hole (m_{HH}^*), averaged light hole (m_{LH}^*), and spin orbit splitoff effective hole masses (m_{SO}) in some cubic III–V semiconductors. ZB = zinc blende

Material	m_{HH}^*/m_0	m_{LH}^*/m_0	m_{SO}/m_0
AlN (ZB)	1.77*	0.35*	0.58*
AlP	0.63*	0.20*	0.29*
AlAs	0.81*	0.16*	0.30*
AlSb	0.9	0.13	0.317*
β -GaN	1.27*	0.21*	0.35*
GaP	0.52	0.17	0.34
GaAs	0.55	0.083	0.165
GaSb	0.37	0.043	0.12
InP	0.69	0.11	0.21
InAs	0.36	0.026	0.14
InSb	0.38	0.014	0.10
* Theoretical			

the splitting of the heavy hole and light hole bands at the Γ point of the strained substance can be explained by the shear deformation potentials, b and d . The lattice mobilities of holes are also strongly affected by these potentials. Several experimental data have been reported on the deformation potential values for III–V alloys, e.g., $\text{Al}_x\text{Ga}_{1-x}\text{As}$ [31.12], $\text{GaP}_x\text{As}_{1-x}$ [31.21] and $\text{Al}_x\text{In}_{1-x}\text{As}$ [31.22]. Due to the large scatter in the experimental binary endpoint values, it is very difficult to establish any evolution of the deformation potentials with composition. We list in Table 31.21 the recommended values for the conduction band (a_c) and valence band deformation potentials (a_v , b , d) of some cubic III–V binaries. The deformation potentials for some wurtzite III–V semiconductors are also collected in Table 31.22. Until more precise data become available, we suggest employing the linear interpolation expressions in order to estimate the parameter values of these poorly explored properties.

Table 31.21 Conduction-band (a_c) and valence-band deformation potentials (a_v , b , d) for some cubic III–V binaries. ZB = zinc blende

Binary	Conduction band	Valence band		
	a_c (eV)	a_v (eV)	b (eV)	d (eV)
AlN (ZB)	−11.7*	−5.9*	−1.7*	−4.4*
AlP	−5.54*	3.15*	−1.5*	
AlAs	−5.64*	−2.6*	−2.3*	
AlSb	−6.97*	1.38*	−1.35	−4.3
β -GaN	−21.3*	−13.33*	−2.09*	−1.75*
GaP	−7.14*	1.70*	−1.7	−4.4
GaAs	−11.0	−0.85	−1.85	−5.1
GaSb	−9	0.79*	−2.4	−5.4
InP	−11.4	−0.6	−1.7	−4.3
InAs	−10.2	1.00*	−1.8	−3.6
InSb	−15	0.36*	−2.0	−5.4
* Theoretical				

Table 31.22 Conduction-band (D_i) and valence-band deformation potentials (C_i) for some wurtzite III–V binaries (in eV)

Binary	Conduction band		Valence band							
	D_1	D_2	C_1	$D_1 - C_1$	C_2	$D_2 - C_2$	C_3	C_4	C_5	C_6
AlN	−10.23*	−9.65*	−12.9*		−8.4*		4.5*	−2.2*	−2.6*	−4.1*
α -GaN	−9.47*	−7.17*	−41.4	−3.1	−33.3	−11.2	8.2	−4.1	−4.7	
InN				−4.05*		−6.67*	4.92*	−1.79*		
* Theoretical										

31.7 Optical Properties

31.7.1 The Reststrahlen Region

It should be noted that in homopolar semiconductors like Si and Ge, the fundamental vibration has no dipole moment and is infrared inactive. In heteropolar semiconductors, such as GaAs and InP, the first-order dipole moment gives rise to a very strong absorption band associated with optical modes that have a k vector of essentially zero (i.e., long-wavelength optical phonons). This band is called the reststrahlen band. Below this band, the real part of the dielectric constant asymptotically approaches the static or low-frequency dielectric constant ϵ_s . The optical constant connecting the reststrahlen near-infrared spectral range is called the high-frequency or optical dielectric constant ϵ_∞ . The value of ϵ_∞ is, therefore, measured for frequencies well above the long-wavelength LO phonon frequency but below the fundamental absorption edge.

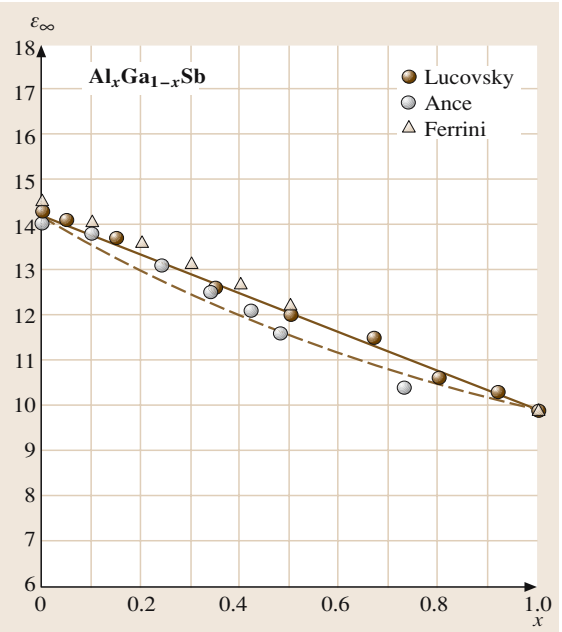


Fig. 31.6 High-frequency dielectric constant ϵ_∞ as a function of x for the $\text{Al}_x\text{Ga}_{1-x}\text{Sb}$ ternary. The experimental data are taken from Lucovsky et al. [31.23] (solid circles), Ance and Mau [31.24] (open circles), and Ferrini et al. [31.25] (open triangles). The solid and dashed lines are, respectively, calculated from (31.1) and (31.6) (ternary) with the binary endpoint values in Table 31.23

Table 31.23 Static (ϵ_s) and high-frequency dielectric constants (ϵ_∞) for some cubic III–V binaries. ZB = zinc blende

Binary	ϵ_s	ϵ_∞
AlN (ZB)	8.16*	4.20
AlP	9.6	7.4
AlAs	10.06	8.16
AlSb	11.21	9.88
β -GaN	9.40*	5.35*
GaP	11.0	8.8
GaAs	12.90	10.86
GaSb	15.5	14.2
InP	12.9	9.9
InAs	14.3	11.6
InSb	17.2	15.3

* Calculated or estimated

Table 31.24 Static (ϵ_s) and high-frequency dielectric constants (ϵ_∞) for some wurtzite III–V binaries

Binary	$E \perp c$		$E \parallel c$	
	ϵ_s	ϵ_∞	ϵ_s	ϵ_∞
AlN	8.3	4.4	8.9	4.8
α -GaN	9.6	5.4	10.6	5.4
InN	13.1*	8.4*	14.4*	8.4*

* Estimated

The general properties of ϵ_s and ϵ_∞ for a specific family of compounds, namely III–V and II–VI compounds, suggest that the dielectric constants in alloy semiconductors could be deduced by using the linear interpolation method [31.26]. The simplest linear interpolation method is to use (31.1), (31.3) or (31.5). The linear interpolation scheme based on the Clausius–Mosotti relation can also be obtained from (31.6). In Fig. 31.6, we show the interpolated ϵ_∞ as a function of x for the $\text{Al}_x\text{Ga}_{1-x}\text{Sb}$ ternary. The solid and dashed lines are, respectively, calculated from (31.1) and (31.6) (ternary). The experimental data are taken from Lucovsky et al. [31.23], Ance and Mau [31.24], and Ferrini et al. [31.25]. The binary endpoint values used in the calculation are listed in Table 31.23. These two methods are found to provide almost the same interpolated values. Table 31.24 also lists the ϵ_s and ϵ_∞ values for some wurtzite III–V binary semiconductors.

The optical spectra observed in the reststrahlen region of alloy semiconductors can be explained by the

following multioscillator model [31.12]:

$$\varepsilon(\omega) = \varepsilon_\infty + \sum_j \frac{S_j \omega_{\text{TO}j}^2}{\omega_{\text{TO}j}^2 - \omega^2 - i\omega\gamma_j}, \quad (31.22)$$

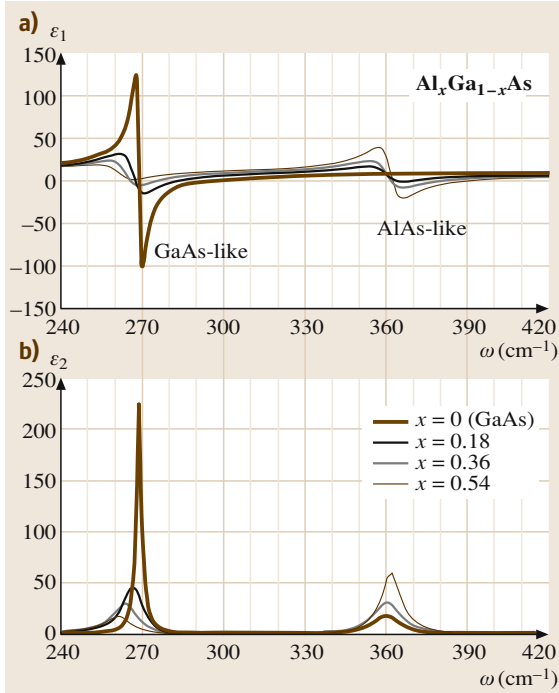


Fig. 31.7 $\varepsilon(\omega)$ spectra in the reststrahlen region of the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary

where $S_j = \varepsilon_\infty (\omega_{\text{LO}j}^2 - \omega_{\text{TO}j}^2)$ is the oscillator strength, $\omega_{\text{TO}j}$ ($\omega_{\text{LO}j}$) is the **TO** (**LO**) phonon frequency, and γ_j is the damping constant of the j -th lattice oscillator. We show in Fig. 31.7, as an example, the optical spectra in the reststrahlen region of the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary. As expected from the two-mode behavior of the long-wavelength optical phonons, the $\varepsilon(\omega)$ spectra of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ exhibit two main optical resonances: GaAs-like and AlAs-like.

31.7.2 The Interband Transition Region

The optical constants in the interband transition regions of semiconductors depend fundamentally on the electronic energy band structure of the semiconductors. The relation between the electronic energy band structure and $\varepsilon_2(E)$ is given by

$$\varepsilon_2(E) = \frac{4e^2\hbar^2}{\pi\mu^2E^2} \int d\mathbf{k} |P_{cv}(\mathbf{k})|^2 \delta[E_c(\mathbf{k}) - E_v(\mathbf{k}) - E], \quad (31.23)$$

where μ is the combined density of states mass, the Dirac δ function represents the spectral joint density of states between the valence-band [$E_v(\mathbf{k})$] and conduction-band states [$E_c(\mathbf{k})$], differing by the energy $E = \hbar\omega$ of the incident light, $P_{cv}(\mathbf{k})$ is the momentum matrix element between the valence-band and conduction-band states, and the integration is performed over the first Brillouin zone. The Kramers–Kronig relations link $\varepsilon_2(E)$ and $\varepsilon_1(E)$ in a manner that means that $\varepsilon_1(E)$ can be calculated at each photon energy if $\varepsilon_2(E)$ is known explicitly over the entire photon energy

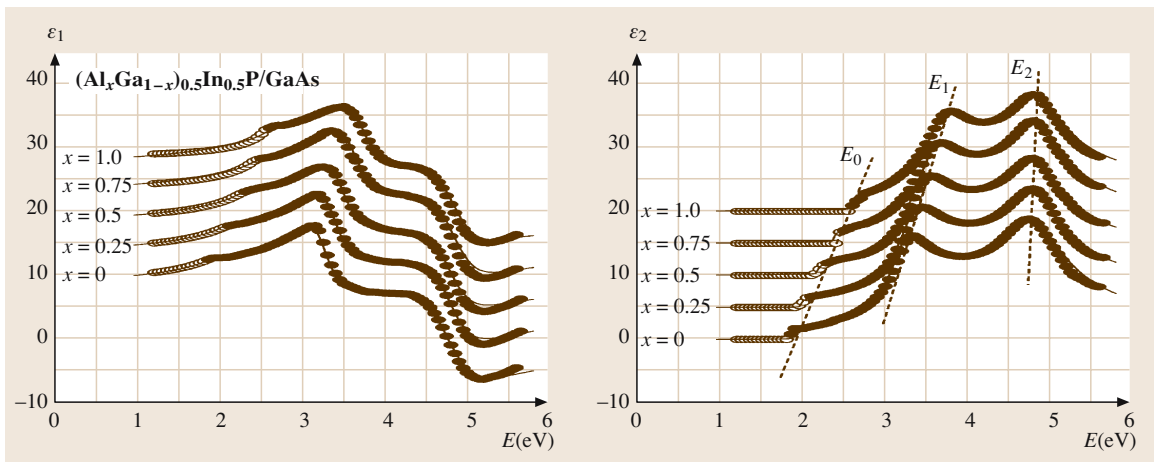


Fig. 31.8 $\varepsilon(E)$ spectra for $\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{P}/\text{GaAs}$ at room temperature. The experimental data are taken from Adachi [31.27]; open and solid circles. The solid lines represent the theoretical fits for the MDF calculation

range, and vice versa. The Kramers–Kronig relations are of fundamental importance in the analysis of optical spectra [31.9].

The refractive indices and absorption coefficients of semiconductors are the basis of many important applications of semiconductors, such as light-emitting diodes, laser diodes and photodetectors. The optical constants of III–V binaries and their ternary and quaternary alloys have been presented in tabular and graphical forms [31.27]. We plot in Fig. 31.8 the $\varepsilon(E)$ spectra for $\text{Al}_x\text{Ga}_y\text{In}_{1-x-y}\text{P}/\text{GaAs}$ taken from tabulation by

Adachi [31.27]; open and solid circles). The solid lines represent the theoretical fits of the model dielectric function (MDF) calculation [31.9]. The three major features of the spectra seen in Fig. 31.8 are the E_0 , E_1 and E_2 structures at ≈ 2 , ≈ 3.5 and ≈ 4.5 eV, respectively. It is found that the E_0 and E_1 structures move to higher energies with increasing x , while the E_2 structure does not do so to any perceptible degree. We can see that the MDF calculation enables us to calculate the optical spectra for optional compositions of alloy semiconductors with good accuracy.

31.8 Carrier Transport Properties

An accurate comparison between experimental mobility and theoretical calculation is of great importance for the determination of a variety of fundamental material parameters and carrier scattering mechanisms. There are various carrier scattering mechanisms in semiconductors, as schematically shown in Fig. 31.9. The effect of the individual scattering mechanisms on the total calculated carrier mobility can be visualized using Matthiessen’s rule:

1/μ_tot = Σ_i 1/μ_i . (31.24)

The total carrier mobility μ_{tot} can then be obtained from the scattering-limited mobilities μ_i of each scattering mechanism. We note that in alloy semiconductors the charged carriers see potential fluctuations as a result of

the composition disorder. This kind of scattering mechanism, so-called alloy scattering, is important in some III–V ternaries and quaternaries. The alloy scattering limited mobility in ternary alloys can be formulated as

μ_al = (√2πeħ^4 N_al α) / (3(m_c^*)^5/2 (kT)^1/2 x(1-x)(ΔU)^2) , (31.25)

where N_{al} is the density of alloy sites, m_c^* is the electron or hole conductivity mass, x and $(1-x)$ are the mole fractions of the binary endpoint materials, and ΔU is the alloy scattering potential. The factor α is caused by the band degeneracy and is given by $\alpha = 1$ for electrons and by $\alpha = [(d^{5/2} + d^3)/(1 + d^{3/2})^2]$ for holes with $d = m_{\text{HH}}/m_{\text{LH}}$, where m_{HH} and m_{LH} are the heavy hole and light hole band masses, respectively [31.12].

Table 31.25 Hall mobilities for electrons (μ_e) and holes (μ_h) obtained at 300 K for relatively pure samples of III–V binaries (in cm^2/Vs)

Binary	μ_e	μ_h
AlP	80	450
AlAs	294	105
AlSb	200	420
α-GaN	1245	370
β-GaN	760	350
GaP	189	140
GaAs	9340	450
GaSb	12 040	1624
InN	3100	
InP	6460	180
InAs	30 000	450
InSb	77 000	1100

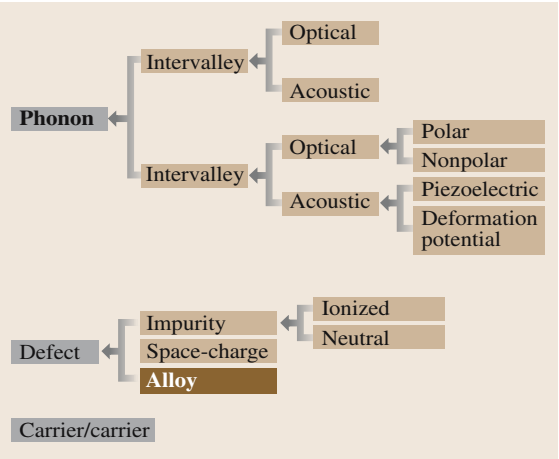


Fig. 31.9 Various possible carrier scattering mechanisms in semiconductor alloys

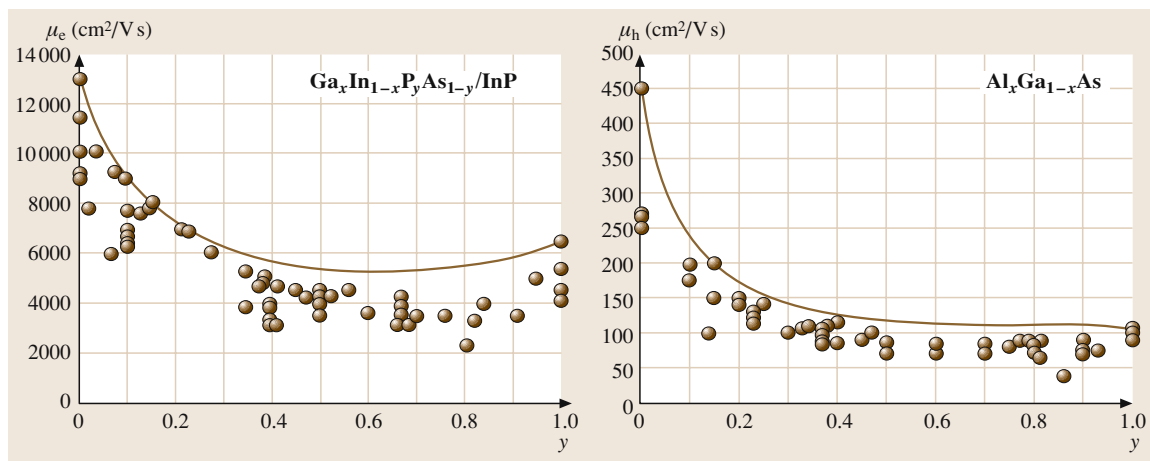


Fig. 31.10 (a) Electron Hall mobility μ_e in the $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}/\text{InP}$ quaternary and (b) the hole Hall mobility μ_h in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ternary, respectively. The experimental data correspond to those for relatively pure samples. The solid lines in (a) and (b) represent the results calculated using (31.26) with $\mu_{\text{al},0} = 3000$ and $50 \text{ cm}^2/\text{Vs}$, respectively

Let us simply express the total carrier mobility μ_{tot} in alloy $\text{A}_x\text{B}_{1-x}\text{C}$ as

$$\frac{1}{\mu_{\text{tot}}(x)} = \frac{1}{x\mu_{\text{tot}}(\text{AC}) + (1-x)\mu_{\text{tot}}(\text{BC})} + \frac{1}{\mu_{\text{al},0}/[x/(1-x)]} \quad (31.26)$$

The first term in (31.26) comes from the linear interpolation scheme and the second term accounts for the effects of alloying.

We plot in Figs. 31.10a and 31.10b the electron Hall mobility in the $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}/\text{InP}$ quaternary (μ_e) and the hole Hall mobility in the $\text{Al}_x\text{Ga}_{1-x}\text{As}$

ternary, respectively. The experimental data correspond to those for relatively pure samples [31.28]. The solid lines in Figs. 31.10a,b represent the results calculated using (31.26) with $\mu_{\text{al},0} = 3000$ and $50 \text{ cm}^2/\text{Vs}$, respectively. The corresponding binary end-point values for μ_{tot} are listed in Table 31.25. For $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}/\text{InP}$, we have considered the quaternary to be an alloy of the constituents $\text{Ga}_{0.53}\text{In}_{0.47}\text{As}$ ($y = 0$) and InP ($y = 1.0$) and we have used the value of μ_{tot} ($\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$) = $13\,000 \text{ cm}^2/\text{Vs}$. It is clear that (31.26) can successfully explain the peculiar composition dependence of the carrier mobility in the semiconductor alloys.

References

- 31.1 A. Mascarenhas: *Spontaneous Ordering in Semiconductor Alloys* (Kluwer Academic, New York 2002)
- 31.2 S. Adachi: *Physical Properties of III-V Semiconductor Compounds: InP, InAs, GaAs, GaP, InGaAs, and InGaAsP* (Wiley-Interscience, New York 1992)
- 31.3 S. Adachi: *Handbook on Physical Properties of Semiconductors, III-V Compound Semiconductors*, Vol. 2 (Springer, Berlin, Heidelberg 2004)
- 31.4 W. Martienssen (Ed): *Landolt-Börnstein, Group III/41 Semiconductors, A1α Lattice Parameters* (Springer, Berlin, Heidelberg 2001)
- 31.5 D. Y. Watts, A. F. W. Willoughby: *J. Appl. Phys.* **56**, 1869 (1984)
- 31.6 D. Cáceres, I. Vergara, R. González, E. Monroy, F. Calle, E. Muñoz, F. Omnès: *J. Appl. Phys.* **86**, 6773 (1999)
- 31.7 M. Krieger, H. Sigg, N. Herres, K. Bachem, K. Köhler: *Appl. Phys. Lett.* **66**, 682 (1995)
- 31.8 W. E. Hoke, T. D. Kennedy, A. Torabi: *Appl. Phys. Lett.* **79**, 4160 (2001)
- 31.9 S. Adachi: *Optical Properties of Crystalline and Amorphous Semiconductors: Materials and Fundamental Principles* (Kluwer Academic, Boston 1999)
- 31.10 C. Pickering: *J. Electron. Mater.* **15**, 51 (1986)
- 31.11 D. H. Jaw, Y. T. Cherng, G. B. Stringfellow: *J. Appl. Phys.* **66**, 1965 (1989)

- 31.12 S. Adachi: *GaAs and Related Materials: Bulk Semiconducting and Superlattice Properties* (World Scientific, Singapore 1994)
- 31.13 A. N. N. Sirota, A. M. Antyukhov, V. V. Novikov, V. A. Fedorov: *Sov. Phys. Dokl.* **26**, 701 (1981)
- 31.14 J. L. Pichardo, J. J. Alvarado-Gil, A. Cruz, J. G. Mendoza, G. Torres: *J. Appl. Phys.* **87**, 7740 (2000)
- 31.15 I. Kudman, R. J. Paff: *J. Appl. Phys.* **43**, 3760 (1972)
- 31.16 J. Bąk-Misiuk, H. G. Brühl, W. Paszkowicz, U. Pietsch: *Phys. Stat. Sol. A* **106**, 451 (1988)
- 31.17 S. Adachi: *J. Appl. Phys.* **54**, 1844 (1983)
- 31.18 I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan: *J. Appl. Phys.* **89**, 5815 (2001)
- 31.19 I. A. Buyanova, W. M. Chen, B. Monemar: *MRS Internet J. Nitride Semicond. Res.* **6**, 2 (2001)
- 31.20 D. L. Young, J. F. Geisz, T. J. Coutts: *Appl. Phys. Lett.* **82**, 1236 (2003)
- 31.21 Y. González, G. Armelles, L. González: *J. Appl. Phys.* **76**, 1951 (1994)
- 31.22 L. Pavesi, R. Houdré, P. Giannozzi: *J. Appl. Phys.* **78**, 470 (1995)
- 31.23 G. Lucovsky, K. Y. Cheng, G. L. Pearson: *Phys. Rev. B* **12**, 4135 (1975)
- 31.24 C. Ance, N. Van Mau: *J. Phys. C* **9**, 1565 (1976)
- 31.25 R. Ferrini, M. Galli, G. Guizzetti, M. Patrini, A. Bosacchi, S. Franchi, R. Magnanini: *Phys. Rev. B* **56**, 7549 (1997)
- 31.26 S. Adachi: *J. Appl. Phys.* **53**, 8775 (1982)
- 31.27 S. Adachi: *Optical Constants of Crystalline and Amorphous Semiconductors: Numerical Data and Graphical Information* (Kluwer Academic, Boston 1999)
- 31.28 M. Sotoodeh, A. H. Khalid, A. A. Rezazadeh: *J. Appl. Phys.* **87**, 2890 (2000)

Group III Nitrides

Optical, electrical and mechanical properties of group III nitrides, including of AlN, GaN, InN and their ternary and quaternary compounds are discussed. The driving force for semiconductor nitrides is device applications for emitters and detectors in the visible and ultraviolet (UV) portions of the optical spectrum and high-power amplifiers. Further advances in electronic and optoelectronic devices, which are imperative, require better understanding and precise measurements of the mechanical, thermal, electrical and optical properties of nitride semiconductors. Information available in the literature regarding many of the physical properties of nitrides, especially AlN and InN, is still in the process of evolution, and naturally in the subject of some controversy. This is, in part, a consequence of measurements having been performed on samples of widely varying quality. When possible, these spurious discrepancies have been disregarded. For other materials, too few measurements are available to yield a consensus, in which case the available data are simply reported. The aim of this work is to present the latest

32.1	Crystal Structures of Nitrides	755
32.2	Lattice Parameters of Nitrides	756
32.3	Mechanical Properties of Nitrides	757
32.4	Thermal Properties of Nitrides	761
32.4.1	Thermal Expansion Coefficients...	761
32.4.2	Thermal Conductivity	762
32.4.3	Specific Heat.....	764
32.5	Electrical Properties of Nitrides	766
32.5.1	Low-Field Transport	766
32.5.2	High-Field Transport	775
32.6	Optical Properties of Nitrides	777
32.6.1	Gallium Nitride	778
32.6.2	Aluminium Nitride.....	786
32.6.3	Indium Nitride.....	789
32.7	Properties of Nitride Alloys	791
32.8	Summary and Conclusions	794
	References	795

available data obtained by various experimental observations and theoretical calculations.

During the last three decades, developments in the field of group III nitrides have been spectacular, with major breakthroughs taking place in the 1990s. They have been viewed as a highly promising material system for electronic and optoelectronic applications. As members of the group III nitrides family, AlN, GaN, InN and their alloys are all wide-band-gap materials and can crystallize in both wurtzite and zincblende polytypes. The band gaps of the wurtzite polytypes are direct and range from a possible value of ≈ 0.8 eV for InN, to 3.4 eV for GaN, and to 6.1 eV for AlN. GaN alloyed with AlN and InN may span a continuous range of direct-band-gap energies throughout much of the visible spectrum, well into ultraviolet (UV) wavelengths. This makes the nitride system attractive for optoelectronic applications, such as light-emitting diodes (LEDs), laser diodes (LDs), and UV detectors. Commercialization of bright blue and green LEDs and the possibility of yellow LEDs

paved the way for developing full-color displays. If the three primary-color LEDs, including red, produced by the InGaAlAs system are used in place of incandescent light bulbs in some form of a color-mixing scheme, they would provide not only compactness and longer lifetime, but also lower power consumption for the same luminous flux output. Additional possible applications include use in agriculture as light sources for accelerated photosynthesis, and in health care for diagnosis and treatment. Unlike display and lighting applications, digital information storage and reading require coherent light sources because the diffraction-limited optical storage density increases approximately quadratically with decreasing wavelength. The nitride material system, when adapted to semiconductor lasers in blue and UV wavelengths, offers increased data storage density, possibly as high as 50 Gb per disc with 25 Gb promised soon in the Blu-Ray system. Other equally

attractive applications envisioned include printing and surgery.

When used as UV sensors in jet engines, automobiles, and furnaces (boilers), the devices would allow optimal fuel efficiency and control of effluents for a cleaner environment. Moreover, visible-blind and solar-blind nitride-based photodetectors are also an ideal candidate for a number of applications including early missile-plume detection, UV astronomy, space-to-space communication, and biological effects.

Another area gaining a lot of attention for group III–V nitrides is high-temperature/high-power electronic applications, such as radar, missiles, and satellites as well as in low-cost compact amplifiers for wireless base stations, due to their excellent electron transport properties, including good mobility and high saturated drift velocity. The strongest feature of the group III nitrides compared to other wide-band-gap counterparts is the heterostructure technology that it can support. Quantum wells, modulation-doped heterointerfaces, and heterojunction structures can all be made in this system, giving access to new spectral regions for optical devices and new operational regimes for electronic devices. Other attractive properties of the nitrides include high mechanical and thermal stability, and large piezoelectric constants.

One of the main difficulties that have hindered group III nitride research is the lack of a lattice-matched and thermally compatible substrate material. A wide variety of materials have been studied for nitride epitaxy, including insulating metal oxides, metal nitrides, and other semiconductors. In practice, properties other than the lattice constants and thermal compatibility, including the crystal structure, surface finish, composition, reactivity, chemical, and electrical properties, are also important in determining suitability as a substrate. The substrate employed determines the crystal orientation, polarity, polytype, surface morphology, strain, and the defect concentration of the epitaxial films. The most promising results on more conventional substrates so far have been obtained on sapphire, and SiC. Also coming on the scene are thick freestanding GaN templates. Group III–V nitrides have been grown on Si, NaCl, GaP, InP, SiC, W, ZnO, MgAl₂O₄, TiO₂, and MgO. Other substrates have also been used for nitride growth, including Hf, LiAlO₂ and LiGaO₂. Lateral (lattice constant *a*) mismatched substrates lead to substantial densities of misfit and threading dislocations in broad-area epitaxially deposited GaN on foreign substrate, in the range 10^9 – 10^{10} cm^{−2}. An appropriate surface preparation such as nitridation, deposition of a low-temperature

(LT) AlN or GaN buffer layer, selective epitaxy followed by a type of coalescence called lateral epitaxial overgrowth (LEO) or epitaxial lateral overgrowth (ELOG) can reduce dislocation densities down to 10^6 cm^{−2}. However, these numbers are still high compared to extended-defect densities of essentially zero for silicon homoepitaxy, and 10^2 – 10^4 cm^{−2} for gallium arsenide homoepitaxy. Vertical (lattice constant *c*) mismatch creates additional crystalline defects besetting the layers, including inversion domain boundaries and stacking faults. In addition, mismatch of thermal expansion coefficients between the epitaxial films and the substrate induces stress, which can cause crack formation in the film and substrate for thick films during cooling from the deposition temperature. A high density of defects, which increases the laser threshold current, causes reverse leakage currents in junctions, depletes sheet charge-carrier density in heterojunction field-effect transistors, reduces the charge-carrier mobility and thermal conductivity, and is detrimental to device applications and the achievement of their optimal performance. Thus, substrates capable of supporting better-quality epitaxial layers are always needed to realize the full potential of nitride-based devices.

Nearly every major crystal-growth technique has been developed, including molecular beam epitaxy (MBE), hydride vapor-phase epitaxy (HVPE), and metalorganic chemical vapor deposition (MOCVD), in relation to nitride semiconductors. Several modifications to the conventional MBE method have been implemented for group III nitride growth: growth with ammonia or hydrazine (the latter is not attractive due to safety reasons and success of ammonia), plasma-assisted MBE (PAMBE), metalorganic MBE (MOMBE), pulsed laser deposition (PLD), etc. Among other methods, radio-frequency (RF) and electron–cyclotron resonance (ECR) plasma sources are the most commonly employed devices to activate the neutral nitrogen species in the MBE environment. Although all of these epitaxial methods contend with problems related to the lack of native GaN substrates, and difficulty with nitrogen incorporation, remarkable progress in the growth of high-quality epitaxial layers of group III nitrides by a variety of methods has been achieved.

Although many applications based on nitride semiconductors has emerged and some of them are commercially available, as discussed throughout this chapter, there are many contradictions in identification of the basic physical properties of these materials. In this respect, they are not yet mature. Additionally, knowledge of the fundamental properties is crucial not only from the

physics point of view but also when understanding and optimizing the device structures for better performance. In this chapter, therefore, we present the updated fundamental properties of GaN, AlN and InN, including structural, mechanical, thermal, electrical, and optical properties. The aim is to assist readers newly entering this field

and other interested researchers in accessing the most-recent available data. The reader is also urged to peruse the following publications for more detail information in several aspects of ongoing research in group III nitrides. These consist of books [32.1,2], edited books and handbooks [32.3–10], and review papers [32.11–29].

32.1 Crystal Structures of Nitrides

The crystal structures shared by the group III nitrides are wurtzite, zincblende, and rocksalt. At ambient conditions, the thermodynamically stable phase is wurtzite for bulk AlN, GaN, and InN. The cohesive energy per bond in wurtzite variety is 2.88 eV (63.5 kcal/mol), 2.20 eV (48.5 kcal/mol), and 1.93 eV (42.5 kcal/mol) for AlN, GaN, and InN, respectively [32.30]. Although the calculated energy difference ΔE_{W-ZB} between wurtzite and zincblende lattice is small (−18.41 meV/atom for AlN, −9.88 meV/atom for GaN, and −11.44 meV/atom for InN) [32.31] the wurtzite form is energetically preferable for all three nitrides compared to zincblende. The wurtzite structure has a hexagonal unit cell with two lattice parameters a and c in the ratio of $c/a = \sqrt{8/3} = 1.633$ and belongs to the space group of $P6_3mc$. The structure is composed of two interpenetrating hexagonal close-packed (hcp) sublattices, each of which consists of one type of atom displaced with respect to each other along the three-fold c -axis by an amount $u = 3/8 = 0.375$ in fractional coordinates. Each sublattice includes four atoms per unit cell and every atom of one kind (group III atom) is surrounded by four atoms of the other kind (nitrogen), or vice versa, these being coordinated at the edges of a tetrahedron. For actual nitrides, the wurtzite structure deviates from the ideal arrangement by changing the c/a ratio or the u value [32.31]. It should be pointed out that a strong correlation exists between the c/a ratio and the u parameter; when the c/a ratio decreases, the u parameter increases in such a way that those four tetrahedral distances remain nearly constant through a distortion of tetrahedral angles due to long-range polar interactions. These two slightly different bond lengths will be equal if the following relation holds;

$$u = \left(\frac{1}{3}\right) \left(\frac{a^2}{c^2}\right) + \frac{1}{4}. \quad (32.1)$$

Since the c/a ratio also correlates with the difference between the electronegativities of the two constituents, components with the greatest differences show largest departure from the ideal c/a ratio [32.32]. These two

parameters were obtained experimentally by using the four-circle diffractometry technique. For GaN, the c/a ratio and the value of u are measured as 1.627 and 0.377, respectively, which are close to the ideal value [32.33]. AlN deviates significantly from the ideal parameters: $c/a = 1.601$ and $u = 0.382$. Consequently, the interatomic distance and angles differ by 0.01 Å and 3°, respectively. For InN, no reliable data are available due to the lack of single-crystal InN with a suitable size for single-crystal diffractometry measurement.

A phase transition to the rocksalt (NaCl) structure in group III nitrides takes place at very high external pressures. The reason for this is that the reduction of the lattice dimensions causes the inter-ionic Coulomb interaction to favor ionicity over the covalent nature. The structural phase transition was experimentally observed at the following pressure values: 22.9 GPa for AlN [32.34], 52.2 GPa for GaN [32.35], and 12.1 GPa for InN [32.36]. The space-group symmetry of the rocksalt type of structure is $Fm\bar{3}m$, and the structure is six-fold coordinated. However, rocksalt group III nitrides cannot be stabilized by epitaxial growth.

The zincblende structure is metastable and can be stabilized only by heteroepitaxial growth on cubic substrates, such as cubic SiC [32.37], Si [32.38], MgO [32.39], and GaAs [32.40], reflecting topological compatibility to overcome the intrinsic tendency to form the wurtzite phase. In the case of highly mismatched substrates, there is usually a certain amount of zincblende phase of nitrides separated by crystallographic defects from the wurtzite phase. The symmetry of the zincblende structure is given by the space group $F\bar{4}3m$ and composed of two interpenetrating face-centered cubic (fcc) sublattices shifted by one quarter of a body diagonal. There are four atoms per unit cell and every atom of one type (group III nitrides) is tetrahedrally coordinated with four atoms of other type (nitrogen), and vice versa. The overall equivalent bond length is about 1.623 Å for zincblende structures.

Because of the tetrahedral coordination of wurtzite and zincblende structures, the four nearest neighbors and

twelve next-nearest neighbors have the same bond distance in both structures. The main difference between these two structures lies in the stacking sequence of close-packed diatomic planes. The wurtzite structure consists of triangularly arranged alternating biatomic close-packed (0001) planes, for example Ga and N pairs, thus the stacking sequence of the (0001) plane is AaBbAaBb in the (0001) direction. In contrast, the zincblende structure consists of triangularly arranged atoms in the close-packed (111) planes along the (111) direction with a 60° rotation that causes a stacking order of AaBbCcAaBbCc. Small and large letters stand for the the two different kinds of constituents.

Since none of the three structures described above possess inversion symmetry, the crystal exhibits crys-

tallographic polarity; close-packed (111) planes in the zincblende and rocksalt structures and the corresponding (0001) basal planes in the wurtzite structure differ from the ($\bar{1}\bar{1}\bar{1}$) and (000 $\bar{1}$) planes, respectively. In general, group III (Al, Ga, or In)-terminated planes are denoted as (0001) A plane (referred to as Ga polarity) and group V (N)-terminated planes are designated as (000 $\bar{1}$) B plane (referred to as N polarity). Many properties of the material also depend on its polarity, for example growth, etching, defect generation and plasticity, spontaneous polarization, and piezoelectricity. In wurtzite nitrides, besides the primary polar plane (0001) and associated direction (0001), which is the most commonly used surface and direction for growth, many other secondary planes and directions exist in the crystal structure.

32.2 Lattice Parameters of Nitrides

Like other semiconductors [32.41–43], the lattice parameters of nitride-based semiconductors depend on the following factors [32.44]:

1. free-electron concentration, acting via the deformation potential of a conduction-band minimum occupied by these electrons,
2. the concentration of foreign atoms and defects, and the difference between their ionic radii and the substituted matrix ion,
3. external strains (for example, those induced by substrate) and
4. temperature.

The lattice parameters of any crystalline materials are commonly and most accurately measured by high-resolution X-ray diffraction (HRXRD) usually at a standard temperature of 21 °C [32.45] by using the Bond method [32.46] for a set of symmetrical and asymmetrical reflections. In ternary compounds, the technique is also used for determining the composition, however, taking the strain into consideration is of crucial issues pertinent to heteroepitaxy. For nitrides, the composition can be determined with an accuracy of about 0.1% or less, down to a mole fraction of about 1%, by taking into account the elastic parameters of all nitrides and lattice parameters of AlN and InN. Since these factors may distort the lattice constants from their intrinsic values, there is wide dispersion in reported values. Table 32.1 shows a comparison of measured and calculated lattice parameters reported by several groups for AlN, GaN, and InN crystallized in the wurtzite structure.

AlN crystal has a molar mass of 20.495 g/mol when it crystallizes in the hexagonal wurtzite structure. The lattice parameters range from 3.110–3.113 Å for the a parameter and from 4.978–4.982 Å for the c parameter as reported. The c/a ratio thus varies between 1.600 and 1.602. The deviation from that of the ideal wurtzite crystal is probably due to lattice stability and ionicity. Although the cubic form of AlN is hard to obtain, several reports suggested the occurrence of a metastable zincblende polytype AlN with a lattice parameter of $a = 4.38\text{Å}$ [32.47], which is consistent with the theoretically estimated value [32.48]. The lattice parameter of a pressure-induced rocksalt phase of AlN is 4.043–4.045 Å at room temperature [32.49, 50].

GaN crystallized in the hexagonal wurtzite (WZ) structure with four atoms per cell and has a molecular weight of 83.7267 g/mol. At room temperature, the lattice parameters of WZ-GaN platelets prepared under high pressure and high temperatures with an electron concentration of $5 \times 10^{19} \text{ cm}^{-3}$ are $a = (3.1890 \pm 0.0003) \text{ Å}$ and $c = (5.1864 \pm 0.0001) \text{ Å}$ [32.44]. For GaN powder, a and c values are in the range 3.1893–3.190, and 5.1851–5.190 Å, respectively. It has been reported that free charge is the dominant factor responsible for expanding the lattice proportional to the deformation potential of the conduction-band minimum and inversely proportional to carrier density and bulk modulus. Point defects such as gallium antisites, nitrogen vacancies, and extended defects, such as threading dislocations, also increase the lattice constant of group III nitrides to a lower extent in the heteroepitax-

Table 32.1 Measured and calculated lattice constants of AlN, GaN and InN

Compound	Sample	a (Å)	c (Å)	Ref.
AlN	Bulk crystal	3.1106	4.9795	[32.53]
	Powder	3.1130	4.9816	[32.54]
	Epitaxial layer on SiC	3.110	4.980	[32.55]
	Pseudopotential LDA	3.06	4.91	[32.56]
	FP-LMTO LDA	3.084	4.948	[32.57]
GaN	Homoepitaxial layers [LFEC (low-free electron concentration)]	3.1885	5.1850	[32.58]
	Homoepitaxial layers [HFEC (high-free electron concentration)]	3.189	5.1864	[32.44]
	Relaxed layer on sapphire	3.1892	5.1850	[32.59]
	Powder	3.1893	5.1851	[32.54]
	Relaxed layer on sapphire	3.1878	5.1854	[32.60]
	GaN substrate	3.1896	5.1855	[32.61]
	Pseudotential LDA	3.162	5.142	[32.56]
	FP-LMTO LDA	3.17	5.13	[32.57]
InN	Powder	3.538	5.703	[32.62]
	Pseudopotential LDA	3.501	5.669	[32.56]
	FP-LMTO LDA	3.53	5.54	[32.57]

LDA: Local density approximation; FP-LMTO: pseudopotential linear muffin-tin orbital

ial layers [32.44]. For the zincblende polytype of GaN, the calculated lattice constant based on the measured Ga–N bond distance in WZ-GaN, is $a = 4.503$ Å while the measured values vary between 4.49 and 4.55 Å, indicating that the calculated result lies within acceptable limits [32.37], Si [32.38] MgO [32.39], and GaAs [32.40]. A high-pressure phase transition from the WZ to the rocksalt structure decreases the lattice constant down to $a_0 = 4.22$ Å in the rocksalt phase [32.51]. This is in agreement with the theoretical result of $a_0 = 4.098$ Å obtained from first-principles non-local pseudopotential calculations [32.52].

Due to the difficulties in synthesis and crystal growth, the number of experimental results concerning the physical properties of InN is quite small, and some have only been measured on non-ideal thin films, typically ordered polycrystalline with crystallites in

the 50–500 nm range. Indium nitride normally crystallizes in the wurtzite (hexagonal) structure, like the other compounds of this family, and has a molecular weight of 128.827 g/mol. The measured lattice parameters using a powder technique are in the range of $a = 3.530$ – 3.548 Å and $c = 5.960$ – 5.704 Å with a consistent c/a ratio of about 1.615 ± 0.008 . The ratio approaches the ideal value of 1.633 in samples having a low density of nitrogen vacancies [32.63]. Recently, Paszkowicz [32.64] reported basal and perpendicular lattice parameters of 3.5378 and 5.7033 Å, respectively, for wurtzite-type InN synthesized using a microwave plasma source of nitrogen, having a c/a ratio far from the ideal value. The single reported measurement yields a lattice constant of $a_0 = 4.98$ Å in the zincblende (cubic) form InN occurring in films containing both polytypes [32.63].

32.3 Mechanical Properties of Nitrides

The mechanical properties of materials involve various concepts such as hardness, stiffness constants, Young's and bulk modulus, yield strength, etc. However, the precise determination of the mechanical properties of the group III nitrides is hindered due to the lack of

high-quality large single crystals. However, attempts to estimate and measure the mechanical properties of thin and thick (separated from substrate) epitaxial layers and bulk crystal of nitrides have been made repeatedly. It has been claimed that the most precise technique

used to determine the elastic moduli of compound materials is ultrasonic measurement. Unfortunately, this ultrasonic pulse-echo method requires thick single crystalline samples, about 1 cm thick, to enable measurement of the timing of plane-wave acoustic pulses with sufficient resolution, which makes it almost inapplicable to the group III nitrides. As an optical technique, Brillouin scattering allows the determination of the elastic constants and hence of the bulk moduli through the interaction of light with thermal excitation in a material, in particular acoustic phonons in a crystal. Various forms of X-ray diffraction, such as energy dispersive X-ray diffraction (EDX), angular dispersive X-ray diffraction (ADX) and X-ray absorption spectroscopy (XAS) can also be employed to determine the pressure dependence of the lattice parameters. From these, the experimental equation of state (EOS), (a widely used one is Murnaghan's equation of state) and hence directly the bulk modulus, assuming that it has a linear dependence with the pressure P , can be deduced as [32.65]:

$$V = V_0 \left(1 + \frac{B'P}{B} \right)^{-\frac{1}{B'}}, \quad (32.2)$$

where B and V_0 represent the bulk modulus and unit volume at ambient pressure, respectively, and B' is the derivative of B with respect to pressure. X-ray diffraction leads to the determination of the isothermal bulk modulus, whereas Brillouin scattering leads to the adiabatic bulk modulus. Nevertheless in solids other than molecular solids there is no measurable difference between these two thermodynamic quantities. Besides the experimental investigation many theoretical calculations have been performed on structural and mechanical properties of group III nitrides. Most of the calculations are based on density-functional theory within the local density approximation (LDA) using various types of exchange correlation functionals, and either plane-wave expansion for the pseudopotentials or the linear muffin-tin orbital (LMTO) method.

In hexagonal crystals, there exist five independent elastic constants, C_{11} , C_{33} , C_{12} , C_{13} and C_{44} . C_{11} and C_{33} correspond to longitudinal modes along the [1000] and [0001] directions, respectively. C_{44} and $C_{66} = (C_{11} - C_{12})/2$ can be determined from the speed of sound of transverse modes propagating along the [0001] and [1000] directions, respectively. The remaining constant, C_{13} , is present in combination with four other moduli in the velocity of modes propagating in less-symmetrical directions, such as [0011]. The bulk

modulus is related to the elastic constants by [32.66]

$$B = \frac{(C_{11} + C_{12})C_{33} - 2C_{13}^2}{C_{11} + C_{12} + 2C_{33} - 4C_{13}}. \quad (32.3)$$

In the isotropic approximation, the Young's modulus E and shear modulus G can also be evaluated using the relations $E = 3B(1 - 2\nu)$ and $G = E/2(1 + \nu)$, respectively. The term ν is the Poisson's ratio and is given by $\nu = C_{13}/(C_{11} + C_{12})$ [32.67].

The micro- and nanoindentation methods are widely used in the determination of the hardness of group III nitrides over a wide range of size scales and temperature. Hardness measurements are usually carried out on the (0001) surface of the crystal using a conventional pyramidal or spherical diamond tip, or alternatively, with a sharp triangular indenter (Berhovich). Depth-sensing indentation measurements provide complete information on the hardness and pressure-induced phase transformation of semiconductor materials. Table 32.2 shows the measured and calculated mechanical parameters reported by several groups for AlN, GaN and InN crystallized in wurtzite structure.

From the widely scattered experimental results presented in Table 32.2, the quality of the crystals is clearly one of the main problems for the precise determination of the physical properties of the group III nitrides. This is true especially for InN, where no elastic moduli could be measured, due to difficulties in synthesis and crystal growth. The difference between elastic moduli measured with the same technique (Brillouin scattering) in GaN is further proof that the quality and nature (bulk single crystal or epitaxial layer) of the samples is of primary importance. Nevertheless, with the notable exception of InN, group III nitrides can be considered as hard and incompressible material family members. Their elastic and bulk moduli are of the same order of magnitude as those of diamond. The hardness of semiconductors is often suggested to be dependent on the bonding distance or shear modulus. Indeed, the softest material InN has a smaller shear modulus and larger bonding distance (0.214 nm) compared to GaN (0.196 nm) and AlN (0.192 nm). The temperature dependence of the hardness shows that macroscopic dislocation motion and plastic deformation of GaN and AlN may start at around 1100 °C. The yield strength of bulk single-crystal GaN is found to be 100–300 MPa at 900 °C. The yield strength of AlN was deduced to be \approx 300 MPa at 1000 °C [32.68].

Most applications of group III nitrides depend on the high thermal conductivity of the material, and a funda-

Table 32.2 Some mechanical properties of wurtzite AlN, GaN, and InN obtained by several experimental techniques and theoretical calculations. The units are in GPa

Parameters	AlN (GPa)	GaN (GPa)	InN (GPa)
C_{11}	345 ^[a] , 411 ^[b] , 396 ^[c] , 398 ^[d]	296 ^[o] , 390 ^[s] , 377 ^[t] , 370 ^[v] , 373 ^[w] , 367 ^[c] , 396 ^[d]	190 ^[g] , 223 ^[c] , 271 ^[d]
C_{12}	125 ^[a] , 149 ^[b] , 137 ^[c] , 140 ^[d]	120 ^[o] , 145 ^[s] , 160 ^[t] , 145 ^[v] , 141 ^[w] , 135 ^[c] , 144 ^[d]	104 ^[g] , 115 ^[c] , 124 ^[d]
C_{13}	120 ^[a] , 90 ^[b] , 108 ^[c] , 127 ^[d]	158 ^[o] , 106 ^[s] , 114 ^[t] , 106 ^[v] , 80 ^[w] , 103 ^[c] , 100 ^[d]	121 ^[g] , 92 ^[c] , 94 ^[d]
C_{33}	395 ^[a] , 389 ^[b] , 373 ^[c] , 382 ^[d]	267 ^[o] , 398 ^[s] , 209 ^[t] , 398 ^[v] , 387 ^[w] , 405 ^[c] , 392 ^[d]	182 ^[g] , 224 ^[c] , 200 ^[d]
C_{44}	118 ^[a] , 125 ^[b] , 116 ^[c] , 96 ^[d]	24 ^[o] , 105 ^[s] , 81 ^[t] , 105 ^[v] , 94 ^[w] , 95 ^[c] , 91 ^[d]	10 ^[g] , 48 ^[c] , 46 ^[d]
Poisson's ratio ν	0.287 ^[f] , 0.216 ^[g]	0.38 ^[f] , 0.372 ^[o]	
Bulk modulus B	201 ^[e] , 210 ^[b] , 208 ^[h] , 160 ^[i] , 207 ^[c] , 218 ^[d]	195 ^[o] , 210 ^[s] , 245 ^[y] , 237 ^[z] , 188 ^[a] , 202 ^[c] , 207 ^[d]	139 ^[g] , 125 ^[z] , 141 ^[c] , 147 ^[d] , 146 ^[b]
dB/dP	5.2 ^[j] , 6.3 ^[k] , 5.7 ^[l] , 3.74 ^[m] , 3.77 ^[n]	4 ^[y] , 4.3 ^[z] , 3.2 ^[a] , 4.5 ^[b] , 2.9 ^[c]	12.7 ^[z] , 3.4 ^[b]
Young's modulus E	308 ^[i] , 295 ^[e] , 374 ^[p]	150 ^[o] , 295 ^[d]	
Shear modulus	154 ^[p] , 131 ^[i] , 117 ^[e]	121 ^[r]	43 ^[g]
Yield strength σ_Y	0.3 at 1000 °C ^[r]	15 ^[d] 0.1–0.2 at 900 °C ^[e]	
Hardness	Micro-hardness: 17.7 ^[r] Nano-hardness: 18.0 ^[r]	Micro-hardness: 10.2 ^[r] Nano-hardness: 18–20 ^[d]	Nano-hardness: 11.2 ^[h]

^[a] Ultrasonic measurement on thin film [32.69]; ^[b] Brillouin scattering on single crystal [32.70]; ^[c] Calculated using pseudopotential LDA [32.71]; ^[d] Calculated using FP-LMTO LDA [32.56]; ^[e] Ultrasonic measurement on thin-film AlN [32.72]; ^[f] {0001}, c-plane calculated [32.73]; ^[g] {112 \bar{b} 0}, r-plane calculated [32.73]; ^[h] ADX on single-crystal AlN [32.74]; ^[i] Ultrasonic measurement on sintered, isotropic, polycrystalline AlN ceramic [32.75]; ^[j] Ultrasonic measurement on sintered, isotropic, polycrystalline AlN ceramic [32.75]; ^[k] ADX on single-crystal AlN [32.74]; ^[l] EDXD on polycrystalline AlN [32.49]; ^[m] Calculated using plane-wave pseudopotential [32.76]; ^[n] Calculated using Keating–Harrison model [32.77]; ^[o] Temperature-dependent X-ray diffraction on polycrystalline GaN [32.78]; ^[p] Hardness measurement on single-crystal AlN [32.67]; ^[r] Hardness measurement on bulk single-crystal AlN [32.67]; ^[s] Brillouin spectroscopy on bulk GaN [32.66]; ^[t] Resonance ultrasound method on GaN plate [32.79]; ^[v] Surface-acoustic-wave measurement on GaN grown on sapphire [32.80]; ^[w] Brillouin spectroscopy on GaN substrate grown by LEO [32.61]; ^[y] X-ray absorption spectroscopy on GaN [32.35]; ^[z] X-ray diffraction on bulk GaN [32.36]; ^[a] EDX on bulk GaN [32.34]; ^[b] Calculated using FP-LMTO [32.56]; ^[c] Calculated using plane-wave pseudopotential [32.52]; ^[d] Nanoindentation on bulk GaN [32.81]; ^[e] Hardness on single-crystal GaN [32.82]; ^[f] {0001}, c-plane using Bond's X-ray method on heteroepitaxially grown GaN [32.59]; ^[g] Temperature-dependent X-ray measurements on powder InN [32.78]; ^[h] Hardness measurement for InN grown on sapphire [32.83] ^[o] {0001}, c-plane estimated from elastic constants [32.78];

mental understanding of the thermal properties requires precise knowledge of the vibrational modes on the single crystal. Infrared reflection and Raman spectroscopies have been employed to derive zone-center and some zone-boundary phonon modes in nitrides. The A_1 and E_1 branches are both Raman- and infrared-active, the E_2 branches are Raman-active only, and the B_1 branches are inactive. The A_1 and E_1 modes are each split into longitudinal optic (LO) and transverse optic (TO) components, giving a total of six Raman peaks. Table 32.3 gives a list of observed zone-center optical-phonon wave numbers along with those calculated from several techniques employed for AlN, GaN, and InN.

The phonon dispersion spectrum of AlN has a total of twelve branches: three acoustic and nine optical. *Perlin* et al. measured the effect of pressure on the Raman shift of a single-crystal AlN sample synthesized at high pressure and high temperature and showed that the pressure dependence of the three observed peaks could be fitted to a quadratic law up to 14 GPa [32.84]. *McNeil* et al. reported the complete set of Raman-active phonon modes of AlN on single crystals grown by the sublimation recondensation method and noted that Raman peaks and widths are influenced by oxygen-related defects [32.70]. Recently, vibrational properties of epitaxial AlN [deposited on silicon and sapphire substrates at ≈ 325 K by ion-beam-assisted deposition (IBAD)]

Table 32.3 Optical phonon frequencies of wurtzite AlN, GaN, and InN at the center of the Brillouin zone in units of cm^{-1}

Symmetry	AlN (cm^{-1})	GaN (cm^{-1})	InN (cm^{-1})
A ₁ -TO	614 ^[a] , 667 ^[b] , 607 ^[c] , 612 ^[d] , 601 ^[e]	533 ^[f] , 531 ^[g] , 544 ^[d] , 534 ^[e] , 533 ^[h]	480 ^[h] , 445 ^[i] , 440 ^[i]
E ₁ -TO	673 ^[a] , 667 ^[b] , 679 ^[d] , 650 ^[e]	561 ^[f] , 558 ^[g] , 566 ^[d] , 556 ^[e] , 559 ^[h]	476 ^[h] , 472 ^[i] , 472 ^[i]
A ₁ -LO	893 ^[a] , 910 ^[b]	735 ^[f] , 733 ^[g] , 737 ^[j]	580 ^[h] , 588 ^[i]
E ₁ -LO	916 ^[a] , 910 ^[b] , 924 ^[c]	743 ^[f] , 740 ^[g] , 745 ^[j]	570 ^[h]
E ₂ -(low)	252 ^[a] , 241 ^[c] , 247 ^[d] , 228 ^[e]	144 ^[f] , 144 ^[g] , 185 ^[d] , 146 ^[e]	87 ^[h] , 104 ^[i]
E ₂ -(high)	660 ^[a] , 660 ^[b] , 665 ^[c] , 672 ^[d] , 638 ^[e]	569 ^[f] , 567 ^[g] , 557 ^[d] , 560 ^[e]	488 ^[h] , 488 ^[i] , 483 ^[i]
B ₁ -(low)	636 ^[d] , 534 ^[e]	526 ^[d] , 335 ^[e]	200 ^[h] , 270 ^[i]
B ₁ -(high)	645 ^[d] , 703 ^[e]	584 ^[d] , 697 ^[e]	540 ^[h] , 530 ^[i]

- [a] Raman scattering on sublimation recondensation AlN [32.70]
[b] Raman scattering on whisker AlN [32.87]
[c] Raman scattering on synthesized AlN by [32.84]
[d] Calculated using first-principle total energy [32.88]
[e] Calculated using pseudopotential LDA [32.89]
[f] Raman scattering on bulk GaN [32.90]
[g] Raman scattering on GaN substrate grown by LEO [32.61]
[h] Raman study on InN grown on sapphire and calculation based on the pairwise interatomic potentials and rigid-ion Coulomb interaction [32.91]
[i] Raman study on polycrystalline and faceted platelets of InN and calculation using FP-LMTO LDA by [32.92]
[j] Raman study on high-quality freestanding GaN templates grown by HVPE [32.93]

have been investigated by *Ribeiro et al.* [32.85]. Raman scattering measurements revealed interesting features and they argued that, due to the extremely weak Raman signal usually exhibited by AlN films, misidentification of some vibration modes can lead to incorrect interpretation of the crystalline quality of AlN films, in which some of the previous mentioned features have been erroneously ascribed [32.86].

The wide spread of studied material has led to some uncertainty in phonon frequencies in GaN, especially of the LO modes. Coupling to plasmons in highly doped material and stress-induced effects due to lattice mismatch with the substrate might play a role in interpretation of the observed phonon frequencies. Moreover, the strong bond in GaN and the light N atoms result in high phonon frequencies that limit the range of observable impurity-related local vibrational modes to even lighter elements at higher frequencies. So far, few reports have appeared for the infrared and Raman modes, which have been associated with local vibrational modes of impurities, dopants, and hydrogen complexes [32.94,95]. The hydrostatic pressure dependence of the zone-center phonon modes has also been determined in n-type bulk

GaN [32.35] and low-residual-doping GaN grown on sapphire [32.96]. The first- and second-order pressure coefficients of the phonon modes have been derived using a polynomial fit.

Raman and infrared spectroscopy studies in InN samples, some grown on (0001) and some on (1102) sapphire substrates, have been undertaken. The infrared (IR) data for A₁(TO) at 448 cm^{-1} and E₁(TO) at 476 cm^{-1} correlate well with the Raman measurements. All six Raman-active modes in the spectra of InN have been observed, with five of them appearing in one InN sample grown on (1102) sapphire. *Kaczmarczyk et al.* [32.97] also studied the first- and second-order Raman-scattering of both hexagonal and cubic InN grown on GaN and GaAs, respectively, covering the acoustic and optical phonon and overtone region. They obtained good agreement with a theoretical model developed by using a modified valence force. The high quality of the samples gives credence to the Raman frequency values as also evidenced by the narrow line widths of all the Raman lines [6.2 cm^{-1} for E₂ (high) and 11.6 cm^{-1} for A₁(LO)]. Details of all the modes are shown in Table 32.3.

32.4 Thermal Properties of Nitrides

32.4.1 Thermal Expansion Coefficients

The lattice parameters of semiconductors are temperature dependent and quantified by thermal expansion coefficients (TEC), which are defined as $\Delta a/a$ or α_a and $\Delta c/c$ or α_c , in and out of plane, respectively. They are dependent on stoichiometry, the presence of extended defects, as well as the free-carrier concentration. As in the case of the lattice parameter, a large scatter in the published data exists for the TEC, particularly for nitrides as they are grown on foreign substrates with different thermal and mechanical properties.

The temperature dependence of the lattice constants a and c , and the thermal expansion coefficients of hexagonal AlN parallel (α_c) and perpendicular (α_a) to the c -axis are shown Fig. 32.1, which can be fitted by the following polynomials within the temperature range $293 < T < 1700$ K;

$$\Delta a/a_0 = -8.679 \times 10^{-2} + 1.929 \times 10^{-4} T + 3.400 \times 10^{-7} T^2 - 7.969 \times 10^{-11} T^3$$

and

$$\Delta c/c_0 = -7.006 \times 10^{-2} + 1.583 \times 10^{-4} T + 2.719 \times 10^{-7} T^2 - 5.834 \times 10^{-11} T^3. \quad (32.4)$$

Using X-ray techniques across a broad temperature range (77–1269 K), it has been noted that the thermal

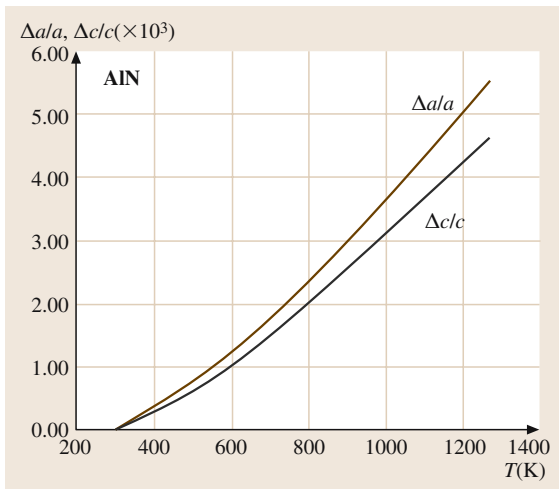


Fig. 32.1 Variation of the thermal expansion coefficient of AlN on temperature, in and out of the c -plane. After [32.98]

expansion of AlN is isotropic with a room-temperature value of $2.56 \times 10^{-6} \text{ K}^{-1}$ [32.100]. The thermal expansion coefficients of AlN measured by Yim and Paff have mean values of $\Delta a/a = 4.2 \times 10^{-6} \text{ K}^{-1}$ and $\Delta c/c = 5.3 \times 10^{-6} \text{ K}^{-1}$ [32.101]. For AlN powder, the expansion coefficients of $2.9 \times 10^{-6} \text{ K}^{-1}$ and $3.4 \times 10^{-6} \text{ K}^{-1}$ has been reported for the a and c parameters, respectively [32.10].

Thermal expansion of single-crystal wurtzite GaN has been studied in the temperature range 300–900 K [32.102] and 80–820 K [32.103]. Maruska and Tietjen reported that the lattice constant a changes linearly with temperature with a mean coefficient of thermal expansion of $\Delta a/a = \alpha_a = 5.59 \times 10^{-6} \text{ K}^{-1}$. Meanwhile, the expansion of the lattice constant

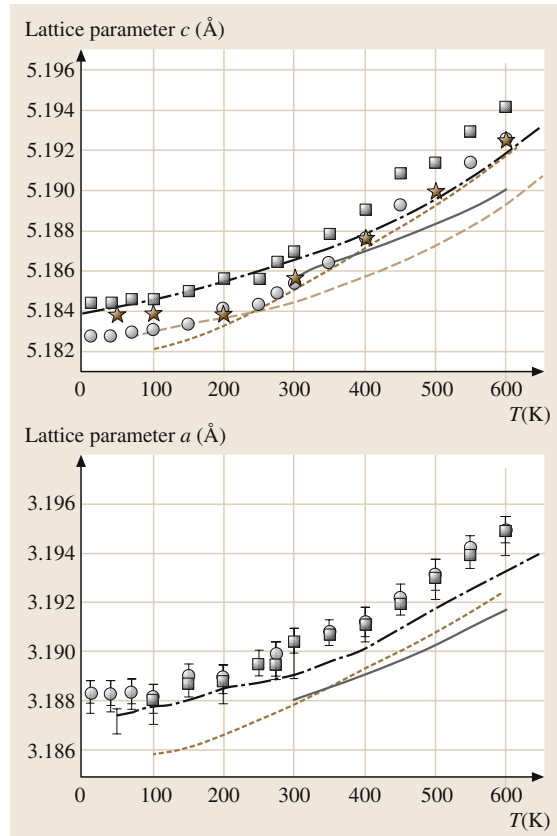


Fig. 32.2 Lattice parameters c and a of a homoepitaxially grown GaN layer, the corresponding substrate and a Mg-doped bulk crystal as a function of temperature in comparison with literature data. After [32.99]

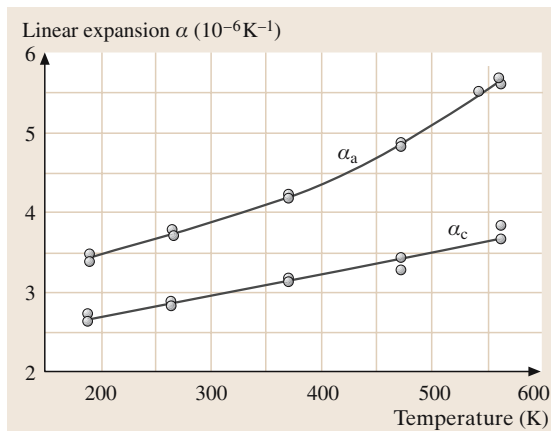


Fig. 32.3 The temperature dependence of linear expansion coefficients in InN. After [32.103]

c shows a superlinear dependence on temperature parallel to the c -axis with a mean coefficients of thermal expansion of $\Delta c/c = \alpha_c = 3.17 \times 10^{-6} \text{ K}^{-1}$ and $7.75 \times 10^{-6} \text{ K}^{-1}$, over the temperature ranges 300–700 K and 700–900 K, respectively. Sheleg and Savastenko reported a thermal expansion coefficient near 600 K of $(4.52 \pm 0.5) \times 10^{-6} \text{ K}^{-1}$ and $(5.25 \pm 0.05) \times 10^{-6} \text{ K}^{-1}$ for the perpendicular and parallel directions, respectively. Leszczynski and Walker reported α_a values of 3.1×10^{-6} and $6.2 \times 10^{-6} \text{ K}^{-1}$, for the temperature ranges of 300–350 K and 700–750 K, respectively [32.104]. The α_c values in the same temperature ranges, were 2.8×10^{-6} and $6.1 \times 10^{-6} \text{ K}^{-1}$, respectively. Being grown on various substrates with different thermal expansion coefficients leads to different dependencies of the lattice parameter on temperature. The temperature dependence of the GaN lattice parameter has been measured for a bulk crystal (grown at high pressure) with a high free-electron concentration ($5 \times 10^{19} \text{ cm}^{-3}$), a slightly strained homoepitaxial layer with a low free-electron concentration (about 10^{17} cm^{-3}), and a heteroepitaxial layer (also with a small electron concentration) on sapphire. Figure 32.2 shows the lattice parameters of various GaN samples along with the theoretical calculation as a function of temperature for comparison.

The linear thermal expansion coefficients measured at five different temperatures between 190 K and 560 K [32.103] indicate that, along both the parallel and perpendicular directions to the c -axis of InN, these coefficients increase with increasing temperature, as shown in Fig. 32.3. The values range from $3.40 \times 10^{-6} \text{ K}^{-1}$ at 190 K to $5.70 \times 10^{-6} \text{ K}^{-1}$ at 560 K and $2.70 \times 10^{-6} \text{ K}^{-1}$

at 190 K to $3.70 \times 10^{-6} \text{ K}^{-1}$ at 560 K for the perpendicular and parallel directions, respectively.

32.4.2 Thermal Conductivity

GaN and other group III nitride semiconductors are considered for high-power/high-temperature electronic and optoelectronic devices where a key issue is thermal dissipation. Consequently, the thermal conductivity (κ), which is the kinetic property determined by the contributions from the vibrational, rotational, and electronic degrees of freedom, is an extremely important material property. The heat transport is predominantly determined by phonon–phonon Umklapp scattering, and phonon scattering by point and extended defects, such as vacancies (including lattice distortions caused by them), impurities such as oxygen, and isotope fluctuations (mass fluctuation). For pure crystals, phonon–phonon scattering is the limiting process. For most groups III nitrides, due to their imperfection, point defects play a significant role for single crystals, .

The thermal conductivity κ of AlN at room temperature has been theoretically estimated to be 3.19 W/cmK for pure AlN single crystals [32.106]. Values of κ meas-

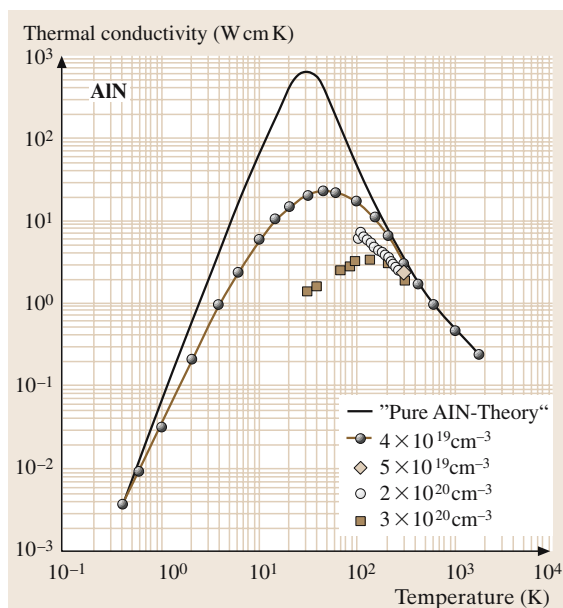


Fig. 32.4 Thermal conductivity of single-crystal AlN. The solid line alone indicates the theory whereas the others represent measurements of AlN with various concentrations of O. The lower the O concentration, the higher the thermal conductivity. After [32.105]

ured at 300 K are 2.5 [32.107] and 2.85 W/cmK [32.100] for single crystals obtained by the sublimation technique, while measurements made on polycrystalline specimens generally yield lower values (reported maximum value of $\kappa = 2.2$ W/cmK) [32.108], due to enhanced phonon O-related point defects and grain-boundary scattering. A more recent prediction of 5.4 W/cmK for the thermal conductivity of AlN is much larger than the value measured by Witek [32.109]. The measured thermal conductivity as a function of temperature in bulk AlN containing some amount of O is plotted in Fig. 32.4. Also shown is a series of samples with estimated concentrations of O, indicating an overall reduction in the thermal conductivity of AlN with O contamination. In the temperature range of interest, where many of the devices would operate, the thermal conductivity in the sample containing the least amount of O assumes a $T^{-1.25}$ dependence. Recently, even higher thermal conductivity values in the range 3.0–3.3 W/cmK were also reported by [32.110] and [32.111] for freestanding and 300–800 μm -thick AlN samples grown by HVPE originally on Si (111) substrates with a 10^8 cm^{-2} dislocation density.

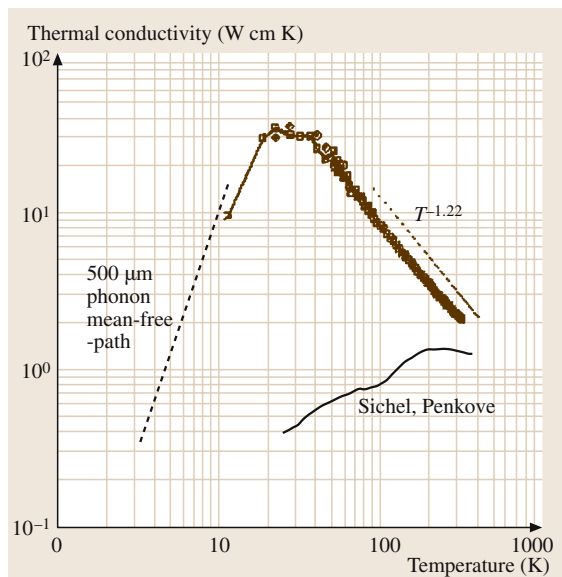


Fig. 32.5 The thermal conductivity of a 200- μm -thick freestanding GaN sample as a function of temperature. The dashed line indicates a calculation using the boundary scattering limit for a phonon mean free path of 500 μm . Also shown is the $T^{-1.22}$ dependence in the range 80–300 K, and earlier results from [32.112] measured using a 400- μm HVPE sample. After [32.105]

The thermal conductivity of GaN layers grown on sapphire substrates by HVPE were measured by [32.112] as a function of temperature (25–360 K) using a heat-flow method. The room-temperature thermal conductivity was measured as $\kappa \cong 1.3$ W/cmK along the c -axis. This value is slightly smaller than the intrinsic value of 1.7 W/cmK for pure GaN predicted by [32.106] and much smaller than the $\kappa \approx 4.10$ W/cmK calculated by [32.109]. More recently, the room-temperature thermal conductivity of GaN was computed as $\kappa = 2.27$ W/cmK, assuming that there is no isotope scattering [32.105]. This prediction is very close to the measured value of 2.3 W/cmK at room temperature, which increases to over 10 W/cmK at 77 K, in high-quality freestanding GaN samples using a steady-state four-probe method. As can be seen in Fig. 32.5, the measured thermal conductivity of GaN in the temperature range 80–300 K has a $T^{-1.22}$ temperature dependence. This slope is typical of pure adamantane crystals below the Debye temperature, indicating acoustic phonon transport where the phonon–phonon scattering is a combination of acoustic–acoustic and acoustic–optic interactions. This temperature dependence strongly suggests that the thermal conductivity depends mainly on intrinsic phonon–phonon scattering and not on phonon–impurity scattering.

A newer method, scanning thermal microscopy (SThM), has been applied to the measurement of room-temperature thermal conductivity on both fully and partially coalesced lateral epitaxial overgrown (LEO) GaN/sapphire (0001) samples. A correlation between low threading-dislocation density and high thermal conductivity values was established. The reduction in the thermal conductivity with increased dislocation density is expected as threading dislocations degrade the velocity of sound and increase phonon scattering in the material. The highest GaN κ values using this method, in the 2.0–2.1 W/cmK range, were found in the regions of the samples that were laterally grown and thus contained a low density of threading dislocations. This compares with a value of 2.3 W/cmK in a freestanding sample measured by the steady-state four-probe method discussed earlier. An explanation for the dramatic increase from $\kappa \approx 1.3$ W/cmK for the early samples to 2.3 W/cmK for the freestanding sample may be related to the extended-defect concentration (D_d) and the differences in background doping. The effect of dislocation density on the thermal conductivity has been calculated [32.113, 114] by showing that κ remains fairly independent of D_d up to some characteristic value D_d^{char} , after which it decreases by about a factor of two for ev-

ery decade increase in D_d . The thermal conductivity has also been correlated to doping levels in hydride vapor-phase epitaxy (HVPE) n-GaN/sapphire (0001) by STHM on two sets of samples [32.115, 116]. In both sets of data the thermal conductivity decreased linearly with $\log n$, where n is the electron concentration, the variation being about a factor of two decrease in κ for every decade increase in n .

InN single crystals of a size suitable for thermal conductivity measurements have not been obtained. The only measurement of thermal conductivity, which has a room temperature value of $\kappa = 0.45 \text{ W/cmK}$, was made on InN ceramics by using the laser-flash method [32.118]. This value is much lower than the thermal conductivity data obtained from the Leibfried–Schloman scaling parameter [32.117]. A value of about $(0.8 \pm 0.2) \text{ W/cmK}$ is predicted assuming that the thermal conductivity is limited by intrinsic phonon–phonon scattering, but this value may be reduced by oxygen contamination and phonon scattering at defects or increased by very high electronic concentrations.

32.4.3 Specific Heat

The specific heat of a semiconductor has contributions from lattice vibrations, free carriers, and point and extended defects. For good-quality semi-insulating crystal, it is only determined by the lattice vibrations. Due to the lack of defect-free crystals of group III nitrides the specific heat measurements are affected by contributions from free carriers and defects, especially at low temperatures. The specific heat C_p of AlN in the temperature interval 298–1800 K has been approximated [32.119] using the expression

$$C_p = 45.94 + 3.347 \times 10^{-3} T - 14.98 \times 10^{-5} T^2 \text{ J/(molK)}. \quad (32.5)$$

For the higher temperature range 1800–2700 K, an approximation [32.120] using the specific heat of $C_p = 51.5 \text{ J/(molK)}$ at $T = 1800 \text{ K}$, and the estimated value of $C_p = 58.6 \text{ J/(molK)}$ at $T = 2700 \text{ K}$, yielded

$$C_p = 37.34 + 7.86 \times 10^{-3} T \text{ J/(molK)}. \quad (32.6)$$

However, since the free electrons (very effective at low temperatures), impurities, defects (inclusive of point defects), and lattice vibrations contribute to the specific heat, these expressions are very simplistic. The Debye expression for the temperature dependence of the specific heat in a solid at constant pressure (C_p) can be

expressed as

$$C_p = 18R \left(\frac{T}{\theta_D} \right)^3 \cdot \int_0^{\theta_D/T} \frac{x^4 e^x}{(e^x - 1)^2} dx, \quad (32.7)$$

where $x_D \equiv \theta_D/T$, and $R = 8.3144 \text{ J/(mol K)}$ is the molar gas constant. The coefficient in front of the term R has been multiplied by 2 to take into account the two constituents making up the group III nitrides. By fitting the measured temperature-dependent heat capacity to the Debye expression, one can obtain the Debye temperature θ_D related to the heat capacity.

The specific heat obtained from the above approximations coupled with the measured values for constant pressure from the literature are shown in Fig. 32.6 along with the calculated specific heat using the Debye equation for Debye temperature values of 800–1100 K with 50 K increments. The best fit between the data and Debye specific-heat expression for insulators indicates a Debye temperature of 1000 K, which is in good agreement with the value of 950 K reported in [32.117].

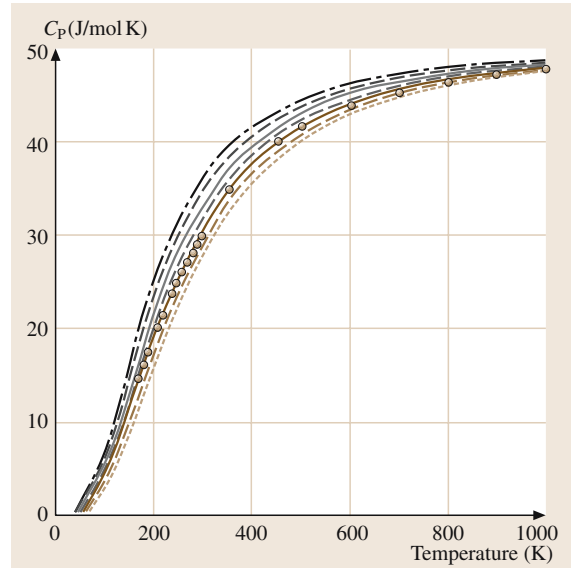


Fig. 32.6 Molar specific heat at constant pressure C_p of AlN versus temperature. Open circles represent the experimental data. The solid lines are calculation based on the Debye model for Debye temperatures θ_D in the range 800–1100 K with 50 K increments. The data can be fitted with the Debye expression for $\theta_D = 1000 \text{ K}$, which compares with the value 950 K reported by Slack et al. The data are taken from [32.117]

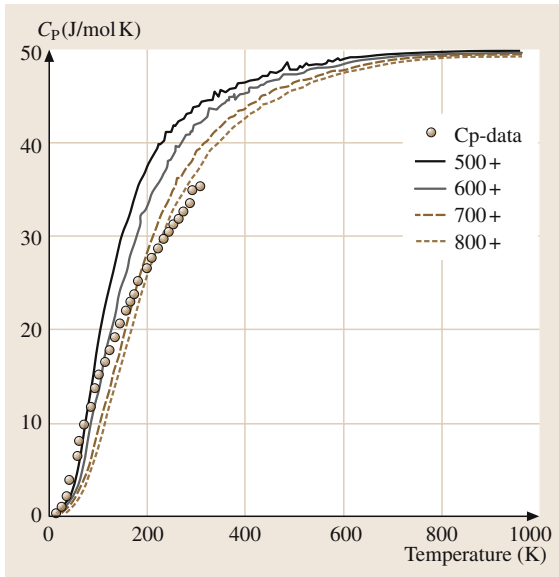


Fig. 32.7 Molar specific heat at constant pressure C_p , of GaN versus temperature. Open circles represent the experimental data. The solid lines are calculation based on the Debye model for Debye temperatures of θ_D of 500, 600, 700, and 800 K. Unfortunately it is difficult to discern a Debye temperature that is effective over a large temperature range because of the high concentration of defects and impurities in GaN used. The data are taken from [32.122]

To a first approximation the temperature dependence of the specific heat of WZ-GaN at constant pressure (C_p) can be expressed by the following phenomenological expression [32.121]

$$C_p = 9.1 + 2.15 \times 10^{-3} T \text{ cal/(molK)}. \quad (32.8)$$

The specific heat of WZ-GaN has been studied in the temperature range 5–60 K [32.122] and also in the temperature range 55–300 K [32.124], and is discussed in [32.123].

The experimental data from [32.122] and [32.124] are plotted in Fig. 32.7. Also shown in the figure is the calculated specific heat using the Debye expression for Debye temperatures of 500, 600, 700 and 800 K. It is clear that the quality of the data and or sample prevent attainment of a good fit between the experimental data and the Debye curve. Consequently, a Debye temperature with sufficient accuracy cannot be determined. It is easier to extract a Debye temperature using data either near very low temperatures or well below the Debye temperature where the specific heat has a simple cubic

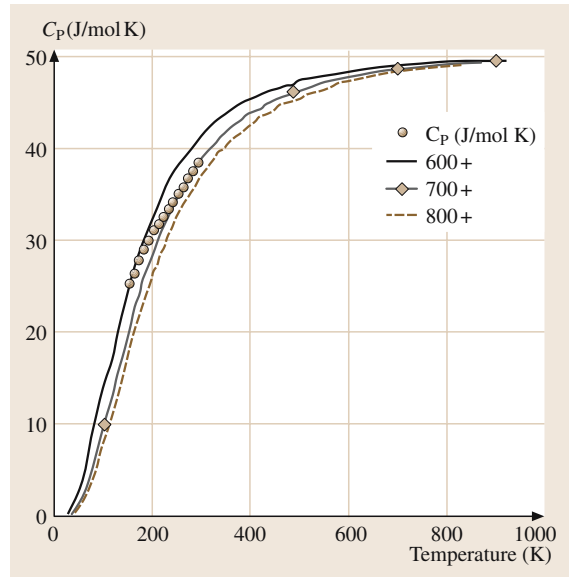


Fig. 32.8 The specific heat of InN with experimental data points, albeit over a small range, and the 600, 700, and 800 K Debye temperature fits. After [32.123]

dependence on temperature [32.125]

$$C_p = 234R \left(\frac{T}{\theta_D} \right)^3. \quad (32.9)$$

Unfortunately, for GaN the samples contain large densities of free carriers and defects, which compromise the application of the Debye specific-heat expression. Consequently, a good fit to the data is not obtained and the Debye temperature extracted in this way is not dependable. Compared to GaN, the Debye temperature obtained in this way for AlN appears more dependable due to a much better fit.

Using the Debye expression for the specific heat and these data, the Debye temperature of InN was obtained as 660 K. The resulting specific-heat curve and the experimental data are plotted in Fig. 32.8. Since the temperature range of these measurements is rather narrow, it is difficult to compare these results and the Debye curve. Good-quality pure InN crystals are extremely difficult to grow and deviations from the Debye curve indicate that the InN samples have significant contributions from non vibrational modes. Using the calculated phonon density of states, the lattice specific heat of InN at constant volume C_v was calculated [32.126] and compared to the experimental values of *Krukowski* et al. [32.123]. This exercise led to the determination of the Debye temperature

θ_D as a function of temperature following the approach of [32.125]. The estimates for θ_D are about

670 K, 580 K, and 370 K at 300 K, 150 K and 0 K, respectively.

32.5 Electrical Properties of Nitrides

GaN and related nitrides being direct and large-band-gap materials lend themselves to a variety of electronic and optoelectronic applications. Advantages associated with a large band gap include higher breakdown voltages, ability to sustain large electric fields, lower noise generation, and high-temperature and high-power operation. Small effective masses in the conduction band minimum lead to reasonably low field mobility, higher satellite-energy separation, and high phonon frequency. Their excellent thermal conductivity, large electrical breakdown fields, and resistance to hostile environments also support the group III nitrides as a material of choice for such applications. The electron transport in semiconductors, including nitrides, can be considered at low and high electric field conditions.

1. At sufficiently low electric fields, the energy gained by the electrons from the applied electric field is small compared to their thermal energy, and therefore the energy distribution of the electrons is unaffected by such a low electric field. Since the scattering rates determining the electron mobility depend on the electron distribution function, electron mobility remains independent of the applied electric field, and Ohm's law is obeyed.
2. When the electric field is increased to a point where the energy gained by electrons from the external field is no longer negligible compared to the thermal energy of the electron, the electron distribution function changes significantly from its equilibrium value. These electrons become hot electrons characterized by an electron temperature larger than the lattice temperature. Furthermore, as the dimensions of the device are shrunk to submicron range, transient transport occurs when there is minimal or no energy loss to the lattice. The transient transport is characterized by the onset of ballistic or velocity-overshoot phenomenon. Since the electron drift velocity is higher than its steady-state value one can design a device operating at higher frequency.

of epitaxial layers. For semiconductor materials, it yields the carrier concentration, its type, and carrier mobility. More specifically, experimental data on Hall measurements over a wide temperature range (4.2–300 K) give information on impurities, imperfections, uniformity, scattering mechanism, etc. The Hall coefficient and resistivity are experimentally determined and then related to the electrical parameters through $R_H = r_H/ne$ and $\mu_H = R_H/\rho$, where n is the free-carrier concentration, e is the unit of electrical charge, μ_H is the Hall mobility, and r_H is the Hall scattering factor. The drift mobility is the average velocity per unit electric field in the limit of zero electric field and is related to the Hall mobility through the Hall scattering factor by $\mu_H = r_H\mu$. The Hall scattering factor depends on the details of the scattering mechanism, which limits the drift velocity. As the carriers travel through a semiconductor, they encounter various scattering mechanisms that govern the carrier mobility in the electronic system. The parameter for characterizing the various scattering mechanisms is the relaxation time τ , which determines the rate of change in electron momentum as it moves about in the semiconductor crystal. Mobility is related to the scattering time by

$$\mu = \frac{q\langle\tau\rangle}{m^*}, \quad (32.10)$$

where μ^* is the electron effective mass, q is the electronic charge, and $\langle\tau\rangle$ is the relaxation time averaged over the energy distribution of electrons. The total relaxation time, τ_T when various scattering mechanisms are operative is given by Matthiessen's rule

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}, \quad (32.11)$$

where i represents each scattering process. The major scattering mechanisms that generally governs the electron transport in group III–V semiconductors is also valid for group III nitrides. They are briefly listed as follows:

1. Ionized-impurity scattering is due to the deflection of free carriers by the long-range Coulomb potential of the charged centers caused by defects or intentionally doped impurities. This can be considered as

32.5.1 Low-Field Transport

The Hall effect is the most widely used technique to measure the transport properties and assess the quality

a local perturbation of the band edge, which affects the electron motion.

2. Polar longitudinal-optical (LO) phonon scattering is caused by the interaction of a moving charge with the electric field induced by electric polarization associated with lattice vibration due to the ionic nature of the bonds in polar semiconductors such as nitrides.
3. Acoustic phonon scattering through the deformation potential arises from the energy change of the band edges induced by strain associated with acoustic phonons, where the scattering rate increases with the wave vectors of the phonons.
4. Piezoelectric scattering arises from the electric fields that are produced by the strain associated with phonons in a crystal without inversion symmetry, particularly in wide-band-gap nitrides.
5. Because of the high density of dislocations and native defects induced by nitrogen vacancies in GaN, dislocation scattering and scattering through nitrogen vacancies has also been considered as a possible scattering mechanism. Dislocation scattering is due to the fact that acceptor centers are introduced along the dislocation line, which capture electrons from the conduction band in an n-type semiconductor. The dislocation lines become negatively charged and a space-charge region is formed around it, which scatters electrons traveling across the dislocations, thus reducing the mobility.

Gallium Nitride

Electron mobility in GaN is one of the most important parameters associated with the material, with a great impact on devices. It has been the subject of intensive studies in recent years from both the experimental and theoretical points of view. Experimental investigation of the temperature-dependent carrier mobility and concentration can be used to determine the fundamental material parameters and understand the carrier scattering mechanism along with an accurate comparison with theory [32.127, 128]. Compared to other group III–V semiconductors, such as GaAs, GaN possesses many unique material and physical properties, as discussed in the previous section. However, the lack of high-quality material, until very recently, prevented detailed investigations of carrier transport. The earlier transport investigations had to cope with poor crystal quality and low carrier mobility, well below predictions [32.129, 130]. Early MBE layers exhibited mobilities as high as $580 \text{ cm}^2/\text{Vs}$ on SiC substrates, which at that time were not as commonly used as in recent times [32.131]. Typically, however, MBE-

grown films produce much lower mobility values of $100\text{--}300 \text{ cm}^2/\text{Vs}$ [32.132]. Different models were used to explain the observed low electron mobilities in GaN, especially at low temperatures. Scattering of electrons at charged dislocation lines [32.132–136] and scattering through elevated levels of point defects [32.137, 138], such as nitrogen vacancies [32.139, 140] were considered as a possible mechanisms responsible for these observations. These scattering mechanisms were investigated by studying the temperature dependence of the carrier concentration and electron mobility. It has been argued that mobility is related to the dislocation density (N_{dis}) and free-carrier concentration (n) via a $\mu_{\text{dis}} \propto \sqrt{n}/N_{\text{dis}}$ relationship [32.135]. At low carrier concentrations ($< 5 \times 10^{17} \text{ cm}^{-3}$), the mobility decreases due to charged dislocation scattering, while at higher carrier concentrations ionized impurities are the dominant mechanism determining the mobility. The temperature dependence of the mobility for samples where dislocations play a dominant role shows that the mobility increases monotonically with temperature, following a $T^{2/3}$ dependence. Electron mobility limited by nitrogen-vacancy scattering was taken into account in n-type GaN grown by MOVPE by Zhu and Sawaki [32.139] and Chen et al. [32.140]. A good fit was obtained between the calculated and experimental results. The estimated mobility shows a $T^{-2/2}$ temperature dependence and it was argued that the measured mobility is dominated by ionized-impurity and dislocation scattering at low temperatures, but polar optical phonon and nitrogen-vacancy scattering at high temperatures.

Hall mobility and electron concentration in undoped GaN were investigated as a function of the thickness of buffer layers and epilayers. Nakamura [32.141] made Hall-effect measurements on undoped GaN layers grown by MOVPE on GaN buffer/sapphire substrates. As the thickness of the buffer layer increased from 100 \AA , the mobility also increased up to a thickness of 200 \AA . At larger thicknesses the mobility began to decrease. The value of the mobility was $600 \text{ cm}^2/\text{Vs}$ at room temperature for a 200 \AA thick layer. The electron concentration was a minimum at 520 nm and increased monotonically as the buffer thickness increased. The mechanism responsible for this observation was not clearly established. Götz et al. [32.142] have studied the effect of the layer thickness on the Hall mobility and electron concentration in unintentionally doped n-type GaN films grown by HVPE on sapphire substrates, which is pretreated with either ZnO or GaCl. They found that the mobility increased and the carrier concentration decreased as the thickness of the epilay-

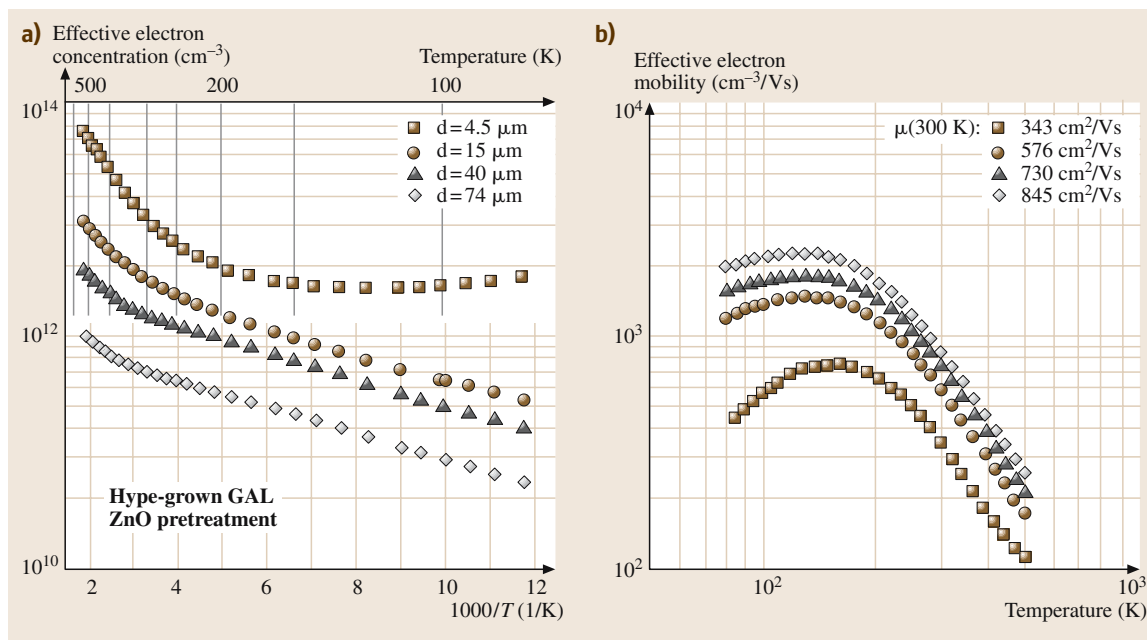


Fig. 32.9a,b Effective electron concentration versus reciprocal temperature (a) and effective electron mobility versus temperature (b), as determined from Hall-effect measurements under the assumption of uniform film properties. After [32.142]

ers increased, as shown in Fig. 32.9. They related this variation to a nonuniform distribution of electrically active defects through the film thickness for both types of films. For GaCl-pretreated sapphire the presence of a highly conductive, 200-nm-thick near-interface layer was assumed to account for the observed phenomena. For ZnO-pretreated sapphire, the Hall-effect data indicated a continuous reduction of the defect density with increasing film thickness.

Since GaN layers are often grown on foreign substrates with very different properties, a degenerate layer forms at the interface caused by extended defects and impurities. Experiments show that, even for thick GaN grown by HVPE, the degenerate interfacial layer makes an important contribution to the Hall conductivity. Look and Molnar [32.143] investigated the Hall effect in the temperature range 10–400 K in HVPE-grown layers on sapphire substrate by assuming a thin, degenerate n-type region at the GaN/sapphire interface. This degenerate interfacial region dominates the electrical properties below 30 K, but also significantly affects those properties even at 400 K, and can cause a second, deeper donor to falsely appear in the analysis. The curve of mobility versus temperature is also affected in that the whole curve is shifted downward from the true, bulk curve. A model

consisting of two layers was constructed to interpret these observations and the result is shown in Fig. 32.10.

Hall mobilities for electron concentrations in the range $\approx 10^{16}$ – 10^{19} cm^{-3} for undoped and intentionally doped (the commonly used donors Si and Ge, which substitute for Ga, are shallow donors with almost identical activation energies for ionization) GaN layers grown by different growth techniques, Fig. 32.11. The most clearly observed trend is that the mobility shows no signs of leveling down to the lowest carrier concentration reported, and the mobility values are practically the same, irrespective of the growth techniques and dopants used, which reflects transport properties inherent to GaN, not to the extrinsic effects. The room-temperature mobility measured by Nakamura et al. [32.144] is $600 \text{ cm}^2/\text{Vs}$ for an electron concentration of $\approx 3 \times 10^{16} \text{ cm}^{-3}$, and decreases slowly with increasing carrier concentration, reaching a value of about $100 \text{ cm}^2/\text{Vs}$ at a carrier concentration of $3 \times 10^{18} \text{ cm}^{-3}$. For an electron concentration of $1 \times 10^{18} \text{ cm}^{-3}$, the mobility is $250 \text{ cm}^2/\text{Vs}$. A higher room-temperature mobility of $845 \text{ cm}^2/\text{Vs}$ at an electron concentration of $\approx 6 \times 10^{16} \text{ cm}^{-3}$ was achieved by Götz et al. [32.142]. In later publications, due to the advent of high-quality samples grown by several

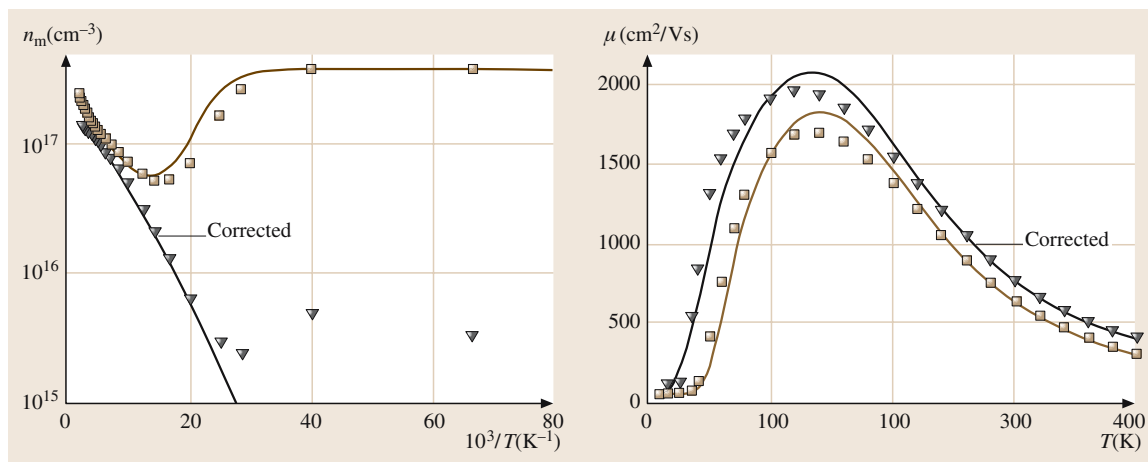


Fig. 32.10 (a) Uncorrected Hall-concentration data (squares) and fit (solid line), and corrected data (triangles) and fit (dashed line) versus inverse temperature. (b) Uncorrected Hall-mobility data (squares) and fit (solid line), and corrected data (triangles) and fit (dashed line) versus temperature. After [32.134]

growth techniques, there is a significant improvement in reported room-temperature mobility values. Nakamura et al. [32.145] and Binari and Dietrich [32.146] were able to obtain 900 cm²/Vs room-temperature mobility at an electron concentration of 3×10^{16} cm⁻³ and 5×10^{16} cm⁻³, respectively. Recently, even higher electron mobilities of 1100 cm²/Vs at room temperature and 1425 cm²/Vs at 273 K were reported in [32.93,147] for a 200- μ m-thick freestanding n-type GaN template grown by HVPE. This achievement was attributed to the excellent crystalline structure of the GaN sample with low levels of compensation and the defect-related scattering. A quantitative comparison with theoretical calculations demonstrates that the one-layer and one-donor conductance model is sufficient to account for the measured data in the entire temperature range without considering any dislocation scattering and any adjustable parameter other than the acceptor concentration. The measured temperature-dependent Hall mobility, carrier concentration and Hall scattering factor is shown in Fig. 32.12 along with the best-fit theoretical calculation based on an iterative solution of the Boltzmann equation. As shown, quantitative agreement with the measured mobility in the entire temperature range was obtained to within about 30%. Heying et al. [32.148] investigated both the morphology and electrical properties of homoepitaxial GaN layers grown by MBE as a function of Ga/N ratio. GaN films grown with higher Ga/N ratios (intermediate regime) showed fewer pits with areas of atomically flat surface, which gives the highest mobility ≈ 1191 cm²/Vs reported so far at room tem-

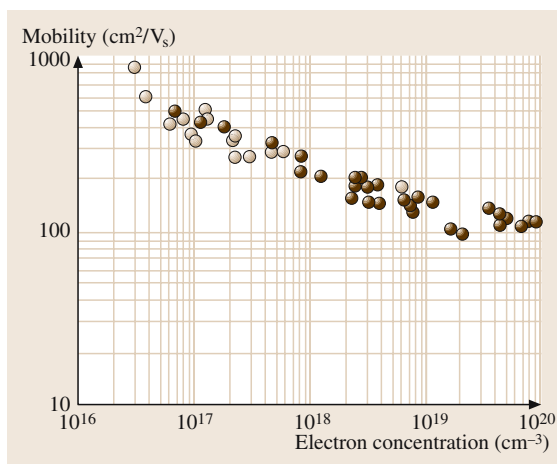


Fig. 32.11 The 300 K Hall mobility versus free electron concentration for GaN from various groups using both MOVPE and MVE. The open circles are from unintentionally doped samples and the solid circles are from samples doped with either Si or Ge. After [32.150]

perature. Koleske et al. [32.149] investigated the effect of the AlN nucleation layer on the transport properties of GaN films grown on 6H- and 4H-SiC substrate. Room-temperature electron mobilities of 876, 884, and 932 cm²/Vs were obtained on 6H-SiC, 4H-SiC, and 3.5° off-axis 6H-SiC substrates. They attributed the observed high electron mobilities to the improved AlN morphology and reduction in screw-dislocation density near the AlN/GaN interface.

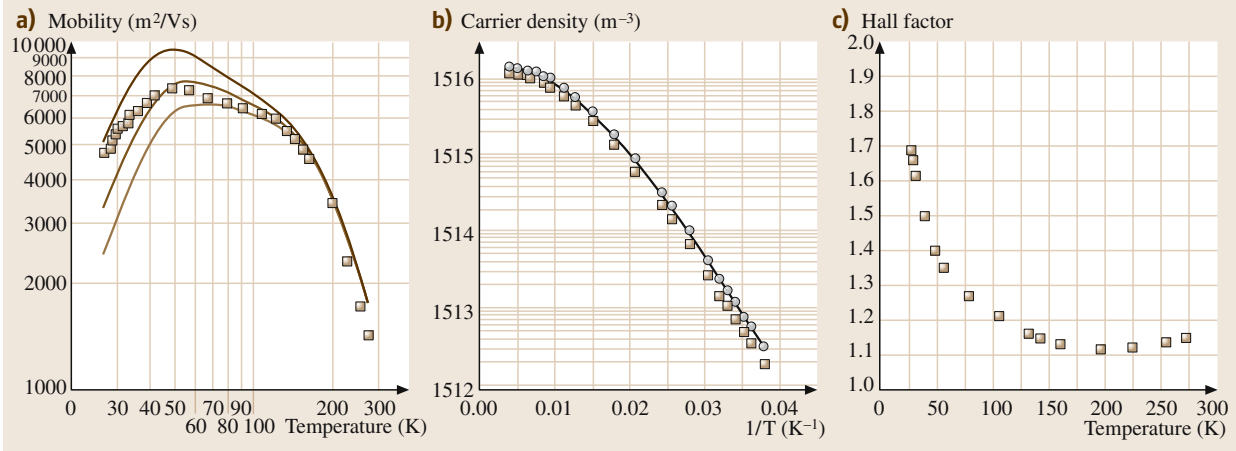


Fig. 32.12 (a) The measured Hall-mobility data (*solid squares*) from the GaN template grown by HVPE as a function of temperature. The *solid line* is the calculated result using $N_a = 2.4 \times 10^{15} \text{ cm}^{-3}$, representing the best fit to the measured results. The *upper* and *lower dotted lines* are the calculated results using $N_a = 1.4 \times 10^{15}$ and $3.4 \times 10^{15} \text{ cm}^{-3}$. (b) The measured Hall densities n_H (*solid squares*) as a function of reciprocal temperature from the GaN template grown by HVPE. The *open circles* represent the carrier density corrected by the Hall factor, $n = n_H r_H$. The *solid line* is the fit to the theoretical expression of charge balance with hole and neutral acceptor densities neglected. (c) Temperature dependence of the calculated Hall factor, r_H . After [32.93]

The thermal activation energy of free carriers has also been extracted from the temperature-dependent Hall concentration and mobility measurements by fitting the simple exponential dependence of the carrier temperature on inverse temperature, by the two-band model and by other theoretical fitting techniques. It could be concluded that the activation energy of n-type GaN free carriers lies in the range 14–36 meV, depending on the extent of screening [32.93, 144, 151, 152]. In the dilute limit, the values are close to 30 meV.

In order to analyze the mobility data, one must understand the scattering processes that dominate mobility at different temperatures. Monte Carlo simulation of the electron velocity in GaN as a function of electric field at different doping concentration and temperature predicted a peak drift velocity of $2 \times 10^7 \text{ cm/s}$ at an electric field of $\approx 1.4 \times 10^5 \text{ V/cm}$ for an electron concentration of 10^{17} cm^{-3} [32.153, 154]. These values show that an electron mobility as high as $900 \text{ cm}^2/\text{Vs}$ could be achieved in the case of uncompensated GaN at room temperature with $\approx 10^{17} \text{ cm}^{-3}$ doping concentration. As discussed above, in the case of high-quality samples with very low compensation, a mobility of even more than $900 \text{ cm}^2/\text{Vs}$ at room temperature with a similar doping concentration has been reported. Albrecht et al. [32.155] calculated the electron mobility for different concentrations of ionized impurities and at different temperatures by using a Monte Carlo simulation technique based on

empirical pseudopotential band-structure calculations. For practical use they have also derived an analytical expression describing the dependence of the mobility on temperature and ionized-impurity concentration as

$$\frac{1}{\mu_e} = a \left(\frac{N_I}{10^{17} \text{ cm}^{-3}} \right) \ln \left(1 + \beta_{CW}^2 \right) \left(\frac{T}{300 \text{ K}} \right)^{-1.5} + b \left(\frac{T}{300 \text{ K}} \right)^{1.5} + c \frac{1}{\exp(\Theta/T) - 1}, \quad (32.12)$$

where

$$\Theta = \frac{\hbar \omega_{LO}}{k_B} = 1065 \text{ K},$$

$$\beta_{WC}^2 = 3.00 \left(\frac{T}{300 \text{ K}} \right)^2 \left(\frac{N_I}{10^{17} \text{ cm}^{-3}} \right),$$

$$N_I = (1 + k_c)/N_D$$

and

$$a = 2.61 \times 10^{-4} \text{ Vs/cm}^2,$$

$$b = 2.90 \times 10^{-4} \text{ Vs/cm}^2 \quad \text{and}$$

$$c = 1.70 \times 10^{-2} \text{ Vs/cm}^2.$$

Here N_D is the ionized donor concentration in cm^{-3} and $k_c = N_A/N_D$ is the compensation ratio. By the mobility calculated using this analytical expression and that calculated by Monte Carlo simulation, a reasonable agreement with a maximum error of 6% is realized

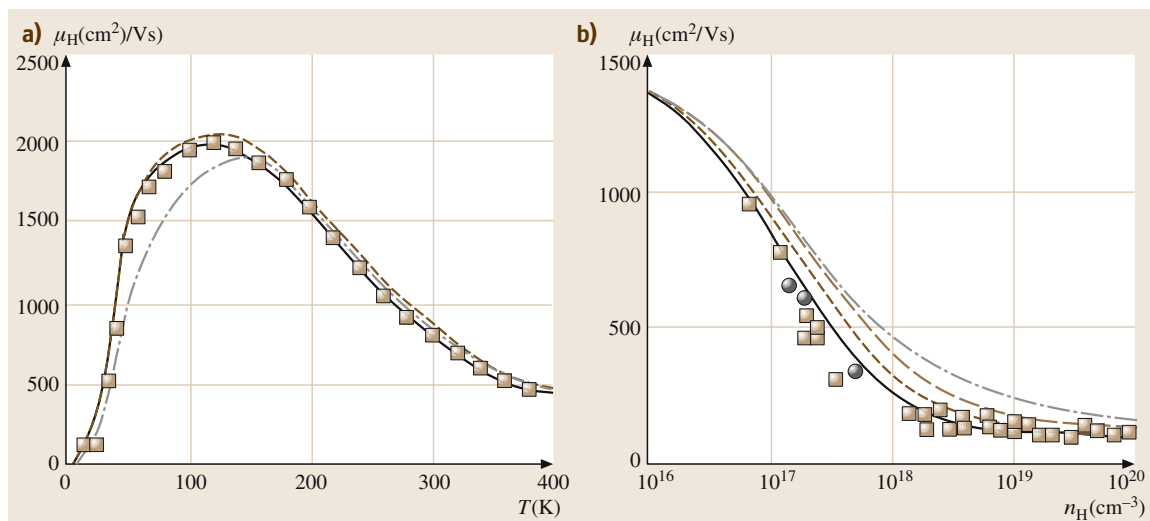
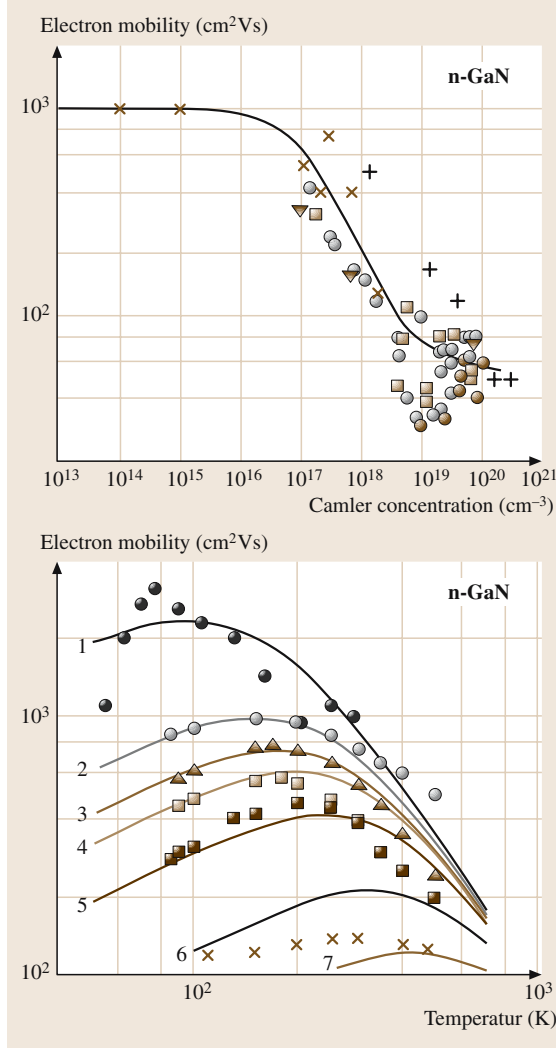


Fig. 32.13 (a) Temperature dependence of the calculated electron Hall mobility. The *dashed curve* shows the calculated mobility using the BH (Brook–Herring) technique. The *dotted curve* shows the calculated mobility using phase-shift analysis with multi-ion screening correction. The *solid curves* show the calculated mobility using phase-shift analysis with multi-ion screening correction and electron–plasmon scattering. (b) Electron concentration dependence of the calculated electron Hall mobility for compensation ratio $r = 0.2$. The *dashed curve* shows the calculated mobility using BH technique. The *dashed–dotted curve* shows the calculated mobility using phase-shift analysis without multi-ion screening correction. The *dotted curve* shows the calculated mobility using phase-shift analysis with multi-ion screening correction. The *solid curve* shows the calculated mobility using phase-shift analysis with multi-ion screening correction and electron–plasmon scattering. Experimental data are taken from different references. After [32.156]

between the temperature range 300–600 K, and ionized donor concentration of 10^{16} – 10^{18} .

The calculation of the low-field electron mobility in GaN has also been carried out by using different calculation techniques. *Chin et al.* [32.129] have used the variational principle to calculate low-field electron mobilities as a function of temperature for carrier concentrations of 10^{16} , 10^{17} , and 10^{18} cm^{-3} with the compensation ratio as a parameter. GaN exhibits maximum mobilities in the range 100–200 K, depending on the electron density and compensation ratio, with lower electron densities peaking at lower temperature. This behavior is related to the interplay of piezoelectric acoustic-phonon scattering at low carrier concentration and ionized-impurity scattering at higher carrier concentrations. The maximum polar-mode optical-phonon scattering-limited room-temperature mobility in GaN is found to be about $1000 \text{ cm}^2/\text{Vs}$. Although a degree of correlation is achieved with the experimental data, there is a disparity, which is attributed to the structural imperfection and overestimated compensation ratio. Typical compensation ratios observed for MOCVD- and MBE-grown films are about 0.3, although

a lower ratio of ≈ 0.24 was reported for HVPE-grown crystals. Compensation reduces the electron mobility in GaN for a given electron concentration. *Rode and Gaskill* [32.130] have used an iterative technique, which takes into account all major scattering mechanisms, for low-field electron mobility in GaN for the dependence of the mobility on the electron concentration, but not on temperature. The result was applied to the Hall-mobility data published by *Kim et al.* [32.157] and a good fit between theory and experiment is demonstrated within 2.5% for the lowest-doped samples with free-electron concentrations of 7.24×10^{17} and $1.74 \times 10^{17} \text{ cm}^{-3}$. However, there is a significant disagreement for the more-heavily-doped samples having much higher free-electron concentrations. The Born approximation applied to the ionized-impurity scattering might be the reason for poor fitting at high electron concentrations. By assuming uncompensated material and carrier freeze out onto donors, the ionization energy was theoretically determined to be about 45 and 57 meV for the best-fitted samples with free-electron concentrations of 7.24×10^{17} and $1.74 \times 10^{17} \text{ cm}^{-3}$, respectively. Recent developments in



the major growth techniques like HVPE, MOCVD, and MBE for GaN and other group III nitride semiconductors have led to the growth of high-quality epitaxial layers [32.158–160], which allowed the comparison of the Hall data with theory to commence. Dhar and Gosh [32.156] have calculated the temperature and doping dependencies of the electron mobility using an iterative technique, in which the scattering mechanisms have been treated beyond the Born approximation. The compensation ratio was used as a parameter with a realistic charge-neutrality condition. They tested their model with the experimental Hall data taken from the literature [32.94, 143–145, 161–168]. Reasonable agreement was achieved between these data points and the

Fig. 32.14 (a) Low-field electron mobility as a function of doping concentration in GaN at room temperature. The curve represents the best approximate equation. The symbols (crosses, open circles, and pluses) are experimental data taken from different references data for wurtzite GaN and zincblende. (b) Temperature dependencies of low-field electron mobility in wurtzite GaN at different values of doping concentration. Curves represent the best approximate equations. 1: $N = 3 \times 10^{16} \text{ cm}^{-3}$, 2: $N = 10^{17} \text{ cm}^{-3}$, 3: $N = 1.5 \times 10^{17} \text{ cm}^{-3}$, 4: $N = 2 \times 10^{17} \text{ cm}^{-3}$, 5: $N = 3.5 \times 10^{17} \text{ cm}^{-3}$, 6: $N = 10^{18} \text{ cm}^{-3}$, 7: $N = 3 \times 10^{18} \text{ cm}^{-3}$. After [32.169]

calculation over a wide range of temperatures and electron concentrations, as shown in Fig. 32.13. They also investigated the effect of the degenerate layer at the GaN/substrate interface to extract reliable experimental values of the bulk electron mobility and concentration.

Besides the numerical simulation techniques mentioned above, recently Mnatsakanov et al. [32.169] derived a simple analytical approximation to describe the temperature and carrier concentration dependencies of the low-field mobility in wide temperature ($50 \leq T \leq 1000 \text{ K}$) and carrier concentration ($10^{14} \leq N \leq 10^{19} \text{ cm}^{-3}$) ranges. At the first step of this model, an adequate approximation of the doping-level dependence of the mobility at room temperature is used on the base of the Goughy–Thomas approximation [32.170]

$$\mu_i(N) = \mu_{\min,i} + \frac{\mu_{\max,i} - \mu_{\min,i}}{1 + \left(\frac{N}{N_{g,i}}\right)^{\gamma_i}}, \quad (32.13)$$

where $i = n, p$ for electrons and holes, respectively, $\mu_{\min,i}$, $\mu_{\max,i}$, $N_{g,i}$ and γ_i are the model parameters dependent on the type of semiconductor materials, and N is the doping concentration. Figure 32.14a shows the comparison between the calculated low-field electron mobility as a function of doping level and some experimental data on the room-temperature electron mobility in GaN. The proposed approximation provides rather good agreement with the experimental data. For the temperature-dependent mobility calculation, the authors derived the following equation by taking into account the main scattering mechanisms:

$$\mu_i(N, T) = \mu_{\max,i}(T_0) \frac{B_i(N) \left(\frac{T}{T_0}\right)^{\beta_i}}{1 + B_i(N) \left(\frac{T}{T_0}\right)^{\alpha_i + \beta_i}}, \quad (32.14)$$

where

$$B_i(N) = \left(\frac{\mu_{\min,i} + \mu_{\max,i} \left(\frac{N_{g,i}}{N} \right)^{\gamma_i}}{\mu_{\max,i} - \mu_{\min,i}} \right) \Big|_{T=T_0}.$$

Figure 32.14b presents the calculated temperature-dependent electron mobility and experimental data reported in the literature and a good agreement is realized. Ionized-impurity scattering is the dominant mechanism at low temperatures, the mobility increases with increasing temperature, at high temperatures mobility is limited by polar optical-phonon scattering, and the mobility decrease with increasing temperature. Finally, many material and physical parameters of GaN were not available for some of the previous simulations where those parameters were treated as adjustable parameters. Needless to say, reliable parameters are required in the calculation of the electron mobility and in the interpretation of experimental results to gain more accuracy.

Aluminium Nitride

Due to the low intrinsic carrier concentration, and the deep native-defect and impurity energy levels (owing to the wide band gap of AlN ≈ 6.2 eV at 300 K), the electrical transport properties of AlN have not been studied extensively and have usually been limited to resistivity measurements. Resistivities in the range $\rho = 10^7 - 10^{13} \Omega\text{cm}$ have been reported for unintentionally doped AlN single crystals [32.171, 172], a value consistent with other reports [32.173–175]. The conductivity exhibited an Arrhenius behavior for all crystals and the activation energies were reported to be 1.4 eV for temperatures of 330 and 400 K and 0.5 eV between 300 and 330 K. Intentional doping of AlN has resulted in both n- and p-type AlN by introducing Hg and Se, respectively. Gorbato and Kamyshev [32.176] obtained the n-type conductivity of polycrystalline AlN with the incorporation Si. Unintentionally doped n-AlN films grown by a modified physical-transport technique by Rutz [32.177] had a resistivity as low as 400 Ωcm . Although the source of the electrons has not been determined, Rutz et al. [32.178] observed an interesting transition in their AlN films in which the resistivity abruptly decreased by two orders of magnitude with an increase in the applied bias. This observation found applications in switchable resistive memory elements that are operated at 20 MHz. It has been concluded by Fara et al. [32.179], who reported the theoretical evidence, based on ab initio density-functional calculations, for acceptors, donors, and native defects in AlN,

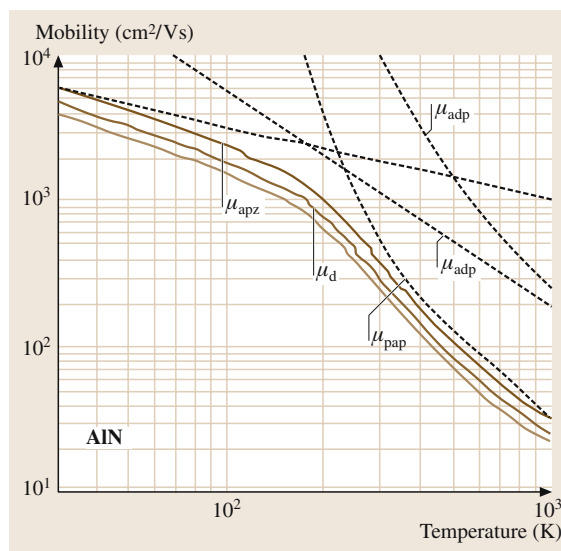


Fig. 32.15 The phonon-limited electron drift mobility in AlN plotted as a function of temperature for $m^* = 0.48$. The upper and lower solid curves are the phonon-limited electron drift mobility calculated using $m^* = 0.42$ and 0.52, respectively. After [32.129]

for the semi-insulating character of AlN that acceptors are deeper (Be ≈ 0.25 , Mg ≈ 0.45) and less soluble in AlN than in GaN, and both the extrinsic donors Si_{Al} and C_{Al}, and the native donor V_N are found to be deep (about 1–3 eV below the conduction). Consequently, doped AlN may in general turn out to be semi-insulating, when attained using Al-rich conditions, in agreement with the known doping difficulties of high-Al-content AlGaN alloys.

As far as the mobility of AlN is concerned, in early studies, Edwards et al. [32.171] and Kawabe et al. [32.172] carried out some Hall measurements in p-type AlN, which produced a very rough estimate of the hole mobility of $\mu_p = 14 \text{ cm}^2/\text{Vs}$ at 290 K. Later work on carbon-doped material has resulted in highly conductive p-type AlN with hole mobilities on the order of 60 cm^2/Vs for carbon doping of 10^{18} cm^{-3} [32.180]. Theoretical calculations to estimate the mobility in AlN have only been made by Chin et al. [32.129]. Since AlN is an insulator with extremely low carrier concentration, only phonon-limited scattering processes were considered in the calculation of the temperature-dependent mobility, as shown in Fig. 32.15. The mobility was found to decline rapidly at high temperatures, with a value of about 2000 cm^2/Vs at 77 K, dropping to 300 cm^2/Vs at 300 K for optical-phonon-limited mobility.

Indium Nitride

InN suffers from the lack of a suitable substrate material, causing a high concentration of extended defects in the films. Sapphire substrate is usually used for InN growth, but the difference in lattice constants between InN and sapphire is as large as 25% and even more than 19% with the other nitrides. A large disparity of the atomic radii of In and N is an additional contributing factor to the difficulty of obtaining InN of good quality. Because of all these factors, the electron mobilities obtained from various films have varied very widely. Electrical properties vary also substantially with the choice of growth techniques. A range of film-deposition techniques, such as reactive-ion sputtering, reactive radio-frequency (RF) magnetron sputtering, metalorganic vapor-phase epitaxy (MOVPE) and MBE, has been used. Table 32.4 gives updated values of electron mobility and concentration in InN films grown by various growth techniques.

The transport properties reported in the literature are mostly based on Hall-effect measurements, assuming the Hall scattering factor to be unity. In majority, electron mobilities often remain relatively poor, despite predicted values as high as 3000 cm²/Vs at room temperature for InN [32.198]. It is widely believed that nitrogen

vacancies lead to large background electron concentrations, which is responsible for the observed low electron mobility. An empirical linear relationship between the electron mobility and electron concentration can be deduced from the table for a series of InN films, although not all films exhibit this type of behavior. A systematic study carried out by Tansley et al. [32.199] indicates that the electron concentration decreases as the nitrogen density is increased in the growth plasma in reactive-ion sputtering. Although the reported time is relatively old, the maximum mobility of 2700 cm²/Vs at an electron concentration of 5 × 10¹⁶ cm⁻³ was reported for RF reactive-ion sputtered growth of InN. Early study of the electron mobility of InN as a function of the growth temperature indicates that the mobility of InN grown by ultrahigh-vacuum electron-cyclotron resonance-radio-frequency magnetron sputtering (UHVECR-RMS) can be as much as four times the mobility of conventionally grown (vacuum-deposited) InN [32.187]. However, more recent work indicates a progressive improvement in electrical properties of InN films grown by vacuum-deposition techniques, including MBE and MOVPE. Values of the electron mobility as high as 1420 cm²/Vs at an electron concentration of 1.4 × 10¹⁸ cm⁻³ were re-

Table 32.4 A compilation of electron mobilities obtained in wide-band-gap InN on different substrates and for various deposition conditions

Growth method	Carrier concentration (cm ⁻³)	Electron mobility (cm/Vs)	Ref.
Reactive ion sputtering	7.0 × 10 ¹⁸	250	[32.181]
Reactive-ion sputtering	2.1 × 10 ¹⁷	470	[32.182]
Reactive-ion sputtering	8.0 × 10 ¹⁶	1300	[32.182]
Reactive-ion sputtering	5.5 × 10 ¹⁶	2700	[32.182]
Reactive-ion sputtering	2.0 × 10 ²⁰	9	[32.183]
RF magnetron sputtering	–	44	[32.184]
Reactive-ion sputter	≈ 10 ²⁰	60	[32.185]
Plasma-assisted MBE	≈ 10 ²⁰	229	[32.63]
ECR-assisted MOMBE	2.0 × 10 ²⁰	100	[32.186]
ECR-assisted reactive-ion sputtering	–	80	[32.187]
Reactive sputter	6.0 × 10 ¹⁸	363	[32.188]
MOVPE	5.0 × 10 ¹⁹	700	[32.159]
Migration-enhanced epitaxy	3.0 × 10 ¹⁸	542	[32.189]
RF MBE	3.0 × 10 ¹⁹	760	[32.190]
MOMBE	8.8 × 10 ¹⁸	500	[32.191]
MBE	2–3 × 10 ¹⁸	800	[32.192]
Reactive-ion sputtering	≈ 10 ¹⁹	306	[32.193]
RF MBE	1.0 × 10 ¹⁹	830	[32.194]
Plasma-assisted MBE	1.6 × 10 ¹⁸	1180	[32.195]
MBE	4 × 10 ¹⁷	2100	[32.196]
Plasma-assisted MBE	1.4 × 10 ¹⁸	1420	[32.197]

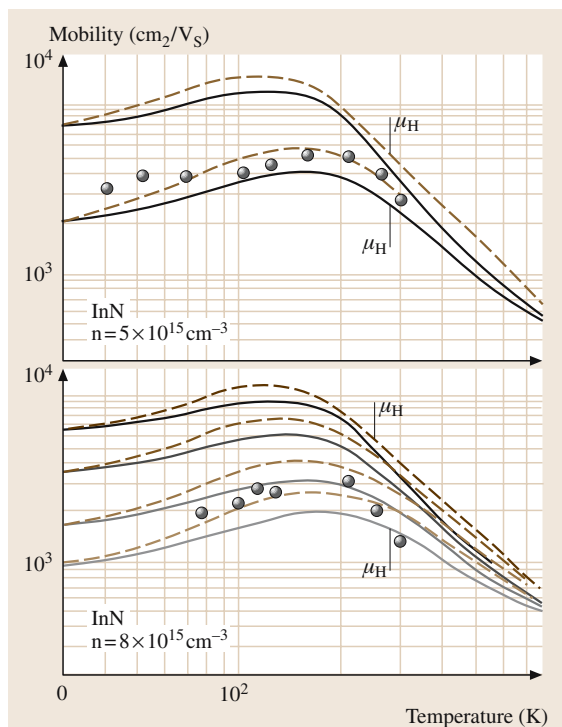


Fig. 32.16a,b The electron drift and Hall mobilities of InN as a function of temperature for (a) $n = 5 \times 10^{16} \text{ cm}^{-3}$ for compensation ratios of 0.00 and 0.60; and (b) $8 \times 10^{16} \text{ cm}^{-3}$ for compensation ratios of 0.00, 0.30, 0.60, and 0.75; all experimental data are from [32.129]

ported for InN layers grown by plasma-assisted MBE using a low-temperature-grown GaN intermediate layer and a low-temperature-grown InN buffer layer. Hall measurements in InN films grown on AlN buffer layers, which are in turn grown on sapphire, indicated an electron mobility of $2100 \text{ cm}^2/\text{Vs}$ with a relatively low electron concentration ($4 \times 10^{17} \text{ cm}^{-3}$) at room temperature in material grown by MBE. Very high inadvertent-donor concentrations ($> 10^{18} \text{ cm}^{-3}$) seem to be one of the major problems for further progress of device applications of InN. O_N and Si_In and possible interstitial H have been proposed to be the likely dominant defects responsible for high electron concentration for state-of-the-art MBE-grown InN, based on their low formation energies [32.200, 201].

The electron mobility in InN has been calculated using the variational principle for a range of temperatures, carrier concentrations, and compensation ratios [32.129]. Figure 32.16 shows the theoretical results with experimental mobility values taken

from Tansley and Foley [32.182]. The calculated peak mobilities are found to be 25 000, 12 000, and $8000 \text{ cm}^2/\text{Vs}$ for 10^{16} , 10^{17} , and 10^{18} cm^{-3} , respectively, at different temperatures (100–200 K), depending on the electron density and the compensation ratio. This is due to the interplay of piezoelectric acoustic-phonon scattering at low concentrations and ionized-impurity scattering at high temperatures. These two mechanisms are the dominant scattering mechanisms below 200 K, while polar-mode optical-phonon scattering is the most significant process above this temperature. The low-concentration limit for room-temperature mobility in uncompensated InN is estimated to be $4400 \text{ cm}^2/\text{Vs}$.

32.5.2 High-Field Transport

Ensemble Monte Carlo simulations have been the popular tools for the theoretical investigation of steady-state electron transport in nitrides. In particular, the steady-state velocity–field characteristics have been determined for AlN [32.203, 204], GaN [32.154, 155, 205–208], and InN [32.209, 210]. These reports show that the

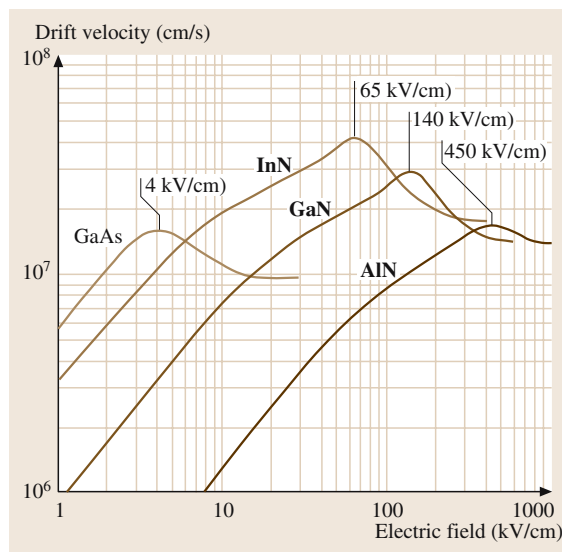


Fig. 32.17 The velocity–field characteristics associated with wurtzite GaN, InN, AlN, and zincblende GaAs. In all cases, we set the temperature to 300 K and the doping concentration to 10^{17} cm^{-3} . The critical fields at which the peak drift velocity is achieved for each velocity–field characteristic are clearly marked: 140 kV/cm for GaN, 65 kV/cm for InN, 450 kV/cm for AlN, and 4 kV/cm for GaAs. After [32.202]

drift velocity initially increases with the applied electric field, reaches a maximum and then decreases with further increases in the field strength. Inter-valley electron transfer plays a dominant role at high electric fields, leading to a strongly inverted electron distribution and to a large negative differential resistance (NDR). The reduction in the drift velocities was attributed to the transfer of electrons from the high-mobility Γ -valley to the low-mobility satellite X-valley. The onset electric field and peak drift velocities, however, show some disparity among the reported calculations due to the variety of degree of approximation and material physical constants used. A typical velocity–field characteristic for bulk group III nitrides at room temperature is shown in Fig. 32.17, along with the well-studied GaAs data

used to test the author's Monte Carlo model. With the doping concentration set to 10^{17} cm^{-3} , InN has the highest steady-state peak drift velocity: $4.2 \times 10^7 \text{ cm/s}$ at an electric field of 65 kV/cm . In the case of GaN and AlN, steady-state peak drift velocities are rather low and occur at larger electric fields: $2.9 \times 10^7 \text{ cm/s}$ at 140 kV/cm for GaN, and $1.7 \times 10^7 \text{ cm/s}$ at 450 kV/cm for AlN.

Another interesting aspect of electron transport is its transient behavior, which is relevant to short-channel devices with dimensions smaller than $0.2 \mu\text{m}$, where a significant overshoot is expected to occur in the electron velocity over the steady-state drift velocity.

Transient electron transport and velocity overshoot in both wurtzite and zincblende GaN, InN, and AlN were studied theoretically by a number of groups. Foutz

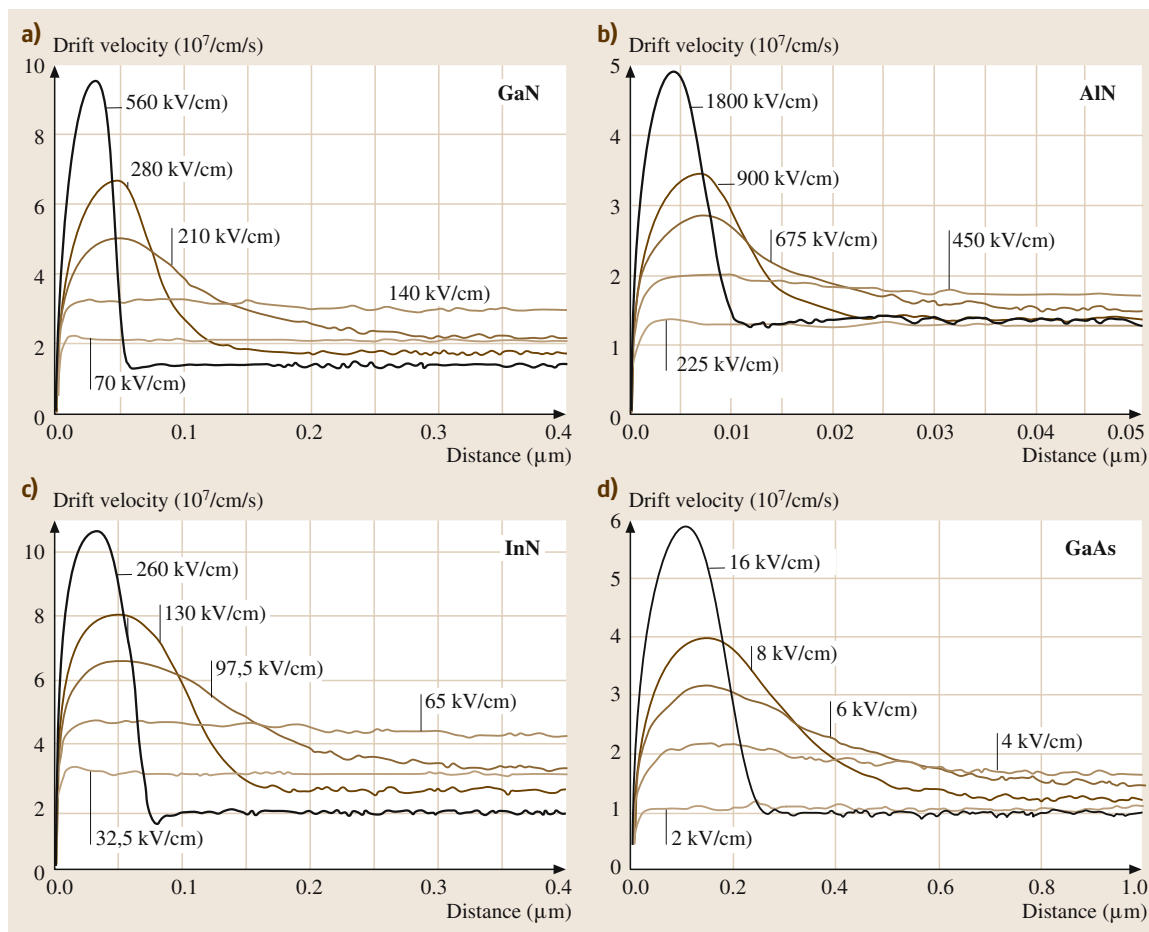


Fig. 32.18a–d The average electron velocity as a function of the displacement for various applied fields for the cases of (a) GaN, (b) InN, (c) AlN, and (d) GaAs. In all cases, we have assumed an initial zero-field electron distribution, a crystal temperature of 300 K, and a doping concentration of 10^{17} cm^{-3} . After [32.202]

et al. [32.202] employed both Monte Carlo simulations and one-dimensional energy-momentum balance techniques. They used a three-valley model for the conduction band by taking the main scattering mechanisms, such as ionized impurity, polar optical-phonon, acoustic-phonon through deformation potential and piezoelectric, and inter-valley scatterings into account. In particular, they examined how electrons, initially in equilibrium, respond to the instant application of a constant electric field. Figure 32.18 shows the average velocity of the electrons in AlN, GaN, and InN as a function of distance. According to their calculation, electron velocity overshoot only occurs when the electric field exceeds a certain critical value unique to each material and it lasts over a certain distance dependent on applied field. These critical fields are points where the highest steady-state peak drift velocities are achieved and being reported as 65 kV/cm, 140 kV/cm, and 450 kV/cm with corresponding peak velocities of 2.9×10^7 cm/s, 1.7×10^7 cm/s, and 1.6×10^7 cm/s for InN, GaN, and AlN, respectively. Among them InN exhibits the highest peak overshoot velocity, on the order of 10^8 cm/s at 260 kV/cm, and the longest overshoot relaxation distance, on the order of 0.8 μ m at 65 kV/cm. To optimize device performance by only minimizing the transit time over a given distance is prevented by a trade-off between the peak overshoot velocity and distance taken to achieve steady state. The upper bound for the cutoff frequency of InN- and GaN-based HFETs (heterojunction field effect transistor) benefits from larger applied fields and accompanying large velocity overshoot when the gate length is less than 0.3 μ m in GaN and 0.6 μ m in InN based devices. However, all measured cutoff frequencies are gate-length-dependent and well below these expectations, indicating that devices operate in the steady-state regime and other effects, such as real-space transfer, should also be considered. On the other hand, there is some controversy in the reports related to the onset of velocity overshoot in nitride semiconductors. For example, *Rodrigues et al.* [32.211] reported that overshoot onsets

at 10 kV/cm in InN, 20 kV/cm in GaN, and 60 kV/cm in AlN by using a theoretical model based on a nonlinear quantum kinetic theory, which compares the relation between the carriers' relaxation rate of momentum and energy.

Experimental investigations of transient transport in group III nitrides are very limited and few results are reported by using different techniques. *Wraback et al.* [32.212] employed a femtosecond time-resolved electroabsorption technique to study transient electron velocity overshoot for transport in the AlGaIn/GaN heterojunction p-i-n photodiode structures. It has been reported that electron velocity overshoot can be observable at electric fields as low as 105 kV/cm. Velocity overshoot increases with electric fields up to ≈ 320 kV/cm with a peak velocity of 7.25×10^7 cm/s relaxing within the first 0.2 ps after photoexcitation. The increase in electron transit time across the device and the decrease in peak velocity overshoot with increasing field beyond 320 kV/cm is attributed to a negative differential resistivity region of the steady-state velocity-field characteristic in this high-field range. *Collazo et al.* [32.213] used another experimental technique based on the measurement of the energy distribution of electrons which were extracted into vacuum through a semitransparent Au electrode, after their transportation through intrinsic AlN heteroepitaxial films using an electron spectrometer. They observed electron velocity overshoot as high as five times the saturation velocity and a transient length of less than 80 nm at a field of 510 kV/cm. In order to design an electronic device that is expected to operate at high power and high frequency, one should consider obtaining benefit from velocity overshoot effect in group III nitrides semiconductor heterojunctions. A systematic investigation of InN, GaN, AlN and their alloys as a function of various parameters in dynamic mode would be very beneficial for the development of higher-performance, next-generation electronic and optoelectronic devices.

32.6 Optical Properties of Nitrides

The optical properties of a semiconductor are connected with both intrinsic and extrinsic effects. Intrinsic optical transitions take place between the electrons in the conduction band and holes in the valance band, including excitonic effects due to the Coulomb interaction. Excitons are classified into free and bound excitons. In high-quality samples with low impurity concentra-

tions, the free exciton can also exhibit excited states, in addition to their ground-state transitions. Extrinsic properties are related to dopants or defects, which usually create discrete electronic states in the band gap, and therefore influence both optical absorption and emission processes. The electronic states of the bound excitons (BEs) depend strongly on the semiconductor material,

in particular its band structure. In theory, excitons could be bound to neutral or charged donors and acceptors. A basic assumption in the description of the principal bound exciton states for neutral donors and acceptors is a dominant coupling of the like particles in the BE states [32.214]. For a shallow neutral donor-bound exciton (DBE), for example, the two electrons in the BE state are assumed to pair off into a two-electron state with zero spin. The additional hole is then assumed to be weakly bound in the net hole-attractive Coulomb potential set up by this bound two-electron aggregate. Similarly, neutral shallow acceptor-bound excitons (ABE) are expected to have a two-hole state derived from the topmost valence band and one electron interaction. These two classes of bound excitons are by far the most important cases for direct-band-gap materials like group III nitrides. Other defect-related transitions could be seen in optical spectra such as free-to-bound (electron–acceptor), bound-to-bound (donor–acceptor) and so-called yellow luminescence. Several experimental techniques are used for the investigation of the optical properties of group III nitrides, including optical absorption, transmission, photoreflection, spectroscopic ellipsometry, photoluminescence, time-resolved photoluminescence, cathodoluminescence, calorimetric spectroscopy, pump-probe spectroscopy, etc. In this section, we will only present some important optical properties of GaN and available data related to AlN and InN.

32.6.1 Gallium Nitride

Free Exciton in GaN

The wurtzite GaN conduction band (Γ_7^c) is mainly constructed from the s state of gallium, whereas the valence band is mainly constructed from the p state of nitrogen. Under the influence of the crystal-field and spin-orbit interactions, the six-fold degenerate Γ_{15} level splits into a highest Γ_9^v , upper Γ_7^v , and lower Γ_7^v levels. The near-band-gap intrinsic absorption and emission spectrum is therefore expected to be dominated by transitions from these three valence bands. The related free-exciton transitions from the conduction band to these three valence bands or vice versa are usually denoted by $A \equiv \Gamma_7^c \leftrightarrow \Gamma_9^v$ (also referred to as the heavy hole), $B \equiv \Gamma_7^c \leftrightarrow \Gamma_7^v$ the upper one (also referred to as the light hole), and $C \equiv \Gamma_7^c \leftrightarrow \Gamma_7^v$ the lower one (also referred to as the crystal-field split band), respectively. The optical properties of these excitonic transitions will be discussed in some detail here. In ideal wurtzite crystals, i.e. strain-free, these three exciton states obey the following selection rules in optical one-photon processes:

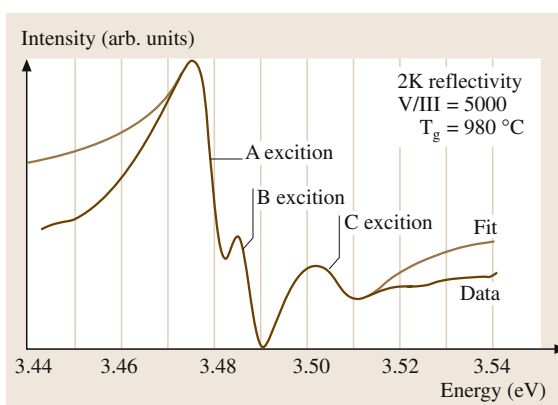


Fig. 32.19 Experimental reflectance spectrum taken at 2 K (solid line) for a 2- μm GaN film grown on sapphire along with a theoretical fit using a damped-oscillator model. After [32.215]

all excitons are allowed in the σ polarization ($E \perp c$ and $k \perp c$ -axis), but the C exciton is quite weak. The C exciton is strongly allowed in the π polarization ($E \parallel c$ and $k \perp c$), however, where the B exciton is weakly observable, the A exciton is forbidden in this geometry. In the α polarization ($E \perp c$ and $k \parallel c$) all three transitions are clearly observable [32.216]. Each of these fundamental exciton states are expected to have a fine structure due both to exciton polariton longitudinal–transverse splitting and the splitting caused by the electron–hole exchange interaction, which are on the order of 1–2 meV [32.217]. Until very recently, this splitting has nearly been impossible to observe due to limitations in the spectroscopic line width of the free exciton on the order of 1 meV in the best sample so far available [32.218]. The optical spectroscopy of the intrinsic excitons can be measured by employing the low-temperature photoluminescence (PL), absorption, and/or derivative technique like photoreflectance (PR) and calorimetric absorption and reflection techniques [32.219, 220]. These measurements pave the way for the determination of exciton binding energies, exciton Bohr radii, the dielectric constant, and with the aid of the quasi-cubic model, spin–orbit and crystal-field parameters. There are several reports in the literature on reflectance studies for thick GaN epilayers [32.216, 221–224] as well as on homoepitaxial layers [32.225, 226] and bulk GaN [32.227]. Monemar et al. [32.220], who have examined numbers of thin and thick layers on various substrates, including homoepitaxial layers on GaN substrates, concluded that the A, B, and C exciton lines in GaN relaxed to an accuracy of ± 2 meV are 3.478, 3.484, and 3.502 eV, respectively, at

2 K. An example of a reflectance spectrum obtained in the α polarization is given in Fig. 32.19 for a 2- μm -thick GaN epilayer on sapphire. The corresponding excitonic transition energies were evaluated with a classical model involving a damped-oscillator transition [32.215].

Calorimetric absorption or reflection is a different experimental technique that has recently been employed to measure the fundamental exciton resonance spectra for a thick GaN epilayer grown on sapphire at 43 mK (Fig. 32.20). The values for the A, B, and C excitons agree with those quoted above for reflection measurement, apart from a small increase of about 2 meV in the case of calorimetric measurement [32.228].

A very powerful technique for studying exciton structure is photoluminescence. Figure 32.21 displays an example of a PL spectrum in the range of fundamental excitons recently taken at 2 K for an approximately strain-free thick GaN epilayer grown on sapphire by HVPE. The PL spectrum is typically dominated by strong emission related to donor-bound excitons (bound excitons will be further discussed later). At higher energies, the $n = 1$ ground states of A and B excitons

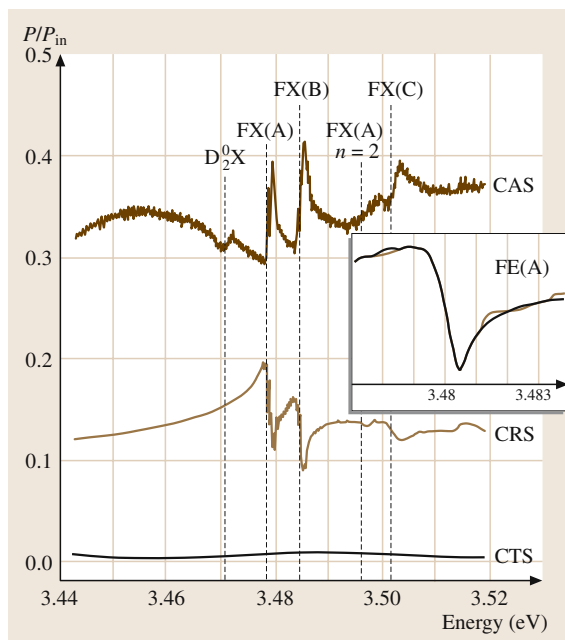


Fig. 32.20 Calorimetric absorption (CAS), reflection (CRS), and transmission (CTS) spectra of the 400- μm -thick HVPE GaN/sapphire layer ($T = 45$ mK). This sample is assumed to be nearly strain-free. In the *inset*, a fit to the reflection spectra of the FE(A) (free exciton) exciton is shown. After [32.228]

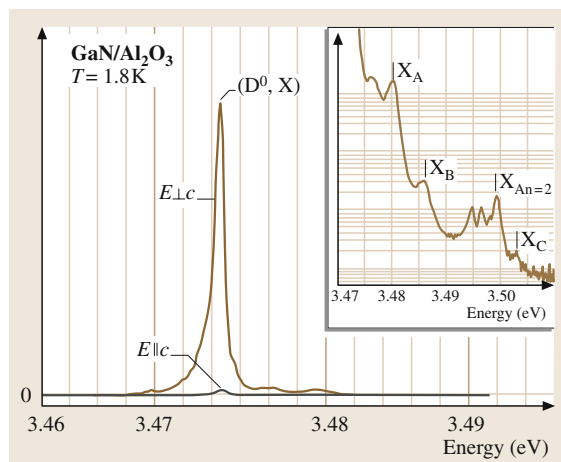


Fig. 32.21 Photoluminescence spectra at 1.8 K for a 400- μm -thick GaN grown on a sapphire substrate. The spectrum is dominated by the donor-bound exciton, but the intrinsic exciton states are also resolved at higher energies (*inset*). After [32.228]

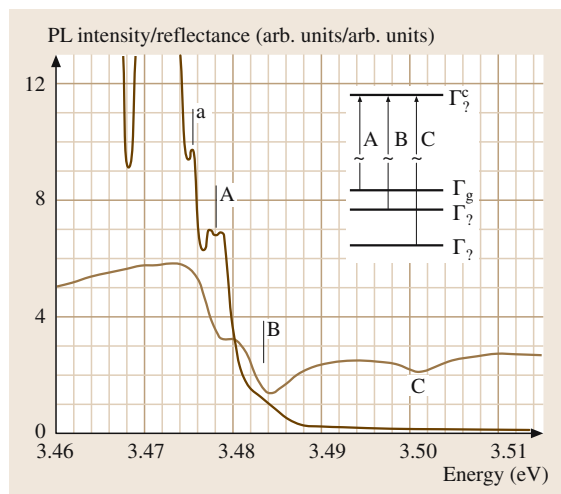


Fig. 32.22 Photoluminescence spectrum (solid line) taken at 4.2 K for a homoepitaxial GaN layer, showing a doublet structure of the A exciton. Corresponding reflectance spectrum for the sample are also shown (dashed line). After [32.229]

are clearly observed with a less-intense C exciton, as well as the $n = 2$ excited state of the A exciton [32.230]. The splitting between the A, B, and C exciton ground states, Δ_{AB} and Δ_{AC} , are (5.5 ± 0.1) and (22.0 ± 0.1) meV. With the aid of these exciton energies, the corresponding values for the splittings

of the three top valance bands have been estimated using a quasi-cubic approximation model. Values of $\Delta_{\text{cr}} \approx 20 \pm 2 \text{ meV}$ for the crystal-field splitting and $\Delta_{\text{so}} \approx 10 \pm 2 \text{ meV}$ for the spin-orbit splitting were reported [32.220]. Edwards et al. [32.171] and Rodina et al. [32.230] have reported different values for the energy splitting parameters: $\Delta_{\text{cr}} \approx \Delta_{\text{so}} \approx 16 \pm 2 \text{ meV}$, and $\Delta_{\text{cr}} \approx 12.3 \pm 0.1 \text{ meV}$, and $\Delta_{\text{so}} \approx 18.5 \pm 0.1 \text{ meV}$, respectively.

Additionally, fine structure of exciton lines was also reported in the energy region near the band edge of GaN using various optical measurements by several groups [32.233–236]. The excitonic spectra taken at 4.2 K for homoepitaxial GaN layers, shown in Fig. 32.22, indicate that the A exciton line splits into two components about 2 meV apart, with the lowest component at about 3.477 eV. A possible interpretation is that these two components actually correspond to recombination from the lower and upper polaron branches. Similarly, Gil et al. [32.237] have also deduced the longitudinal-transverse splitting by re-

flectance line-shape fitting to the data. For each exciton, the longitudinal transverse splitting is calculated as 2.9 meV and 1.8 meV for the A and B excitons, respectively. This compares, within the experimental accuracy, with the splitting of 2.4 and 1.8 meV between the energies of the dips in the PL bands at 3.4894 and 3.4978 eV, and transverse excitons as shown in Fig. 32.22.

The values of the near-band-gap exciton energies are strongly sensitive to built-in strain, which commonly occurs when the GaN is grown on a foreign substrate with heteroepitaxy. GaN grown on sapphire substrates usually experiences a compressive strain, which in turn, increases the band gap, and hence increases in A, B, and C exciton energies and splittings, compared to the case of unstrained bulk GaN, are consistently observed. Upshifts of the A and B excitons by as much as 20 meV have been observed at 2 K, and the C exciton has been found to shift as much as 50 meV [32.222]. An example of the systematic shifts observed in reflectance data for A, B, and C excitons is shown in Fig. 32.23. On the other hand, a different situation is expected to occur for GaN growth on another widely used substrate, SiC, due to the tensile strain, which in turn is expected to lead to a decrease in the overall exciton energies (and the band

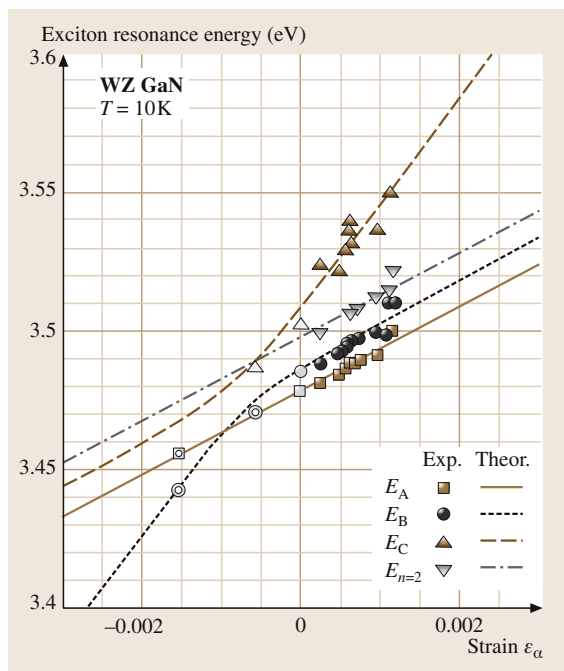


Fig. 32.23 The strain dependence of the free-exciton resonance energies in wurtzite GaN grown on sapphire substrate. The strain was obtained from the measured lattice parameter values for each sample. Theoretical modeling of the strain dependence of exciton energies are also shown. [32.231]

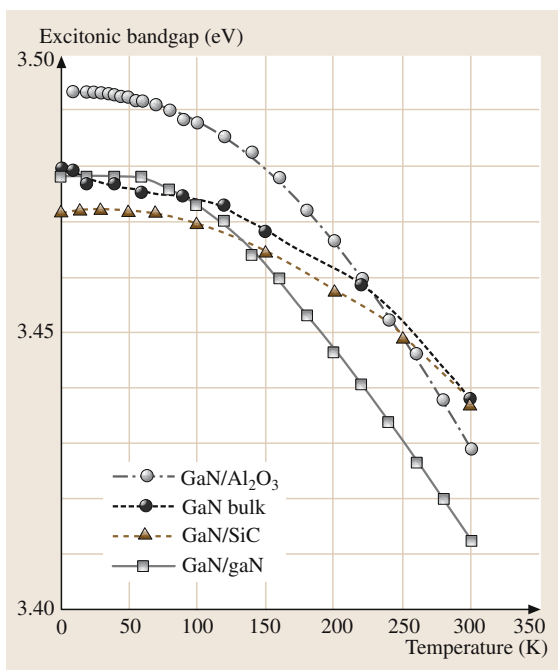


Fig. 32.24 Temperature dependence for the A exciton energy in samples grown on different substrates. After [32.225, 232]

Table 32.5 Reported exciton energies in GaN using different experimental techniques

FX(A)	FX(B)	FX(C)	Substrate	Strain	Growth	Technique	Ref.
3.474	3.481	3.501	Sapphire	Compressive	VPE	Reflectance	[32.214]
3.475	3.481	3.493	Sapphire	Compressive	VPE	PLE	[32.221]
3.480	3.486	3.503	Sapphire	Compressive	HVPE	PL	[32.230]
3.4799	3.4860	3.5025	Sapphire	Compressive	HVPE	Calorimetric reflection	[32.228]
3.4770	3.4865	3.5062	Sapphire	Compressive	MOVPE	Reflectance	[32.215]
3.4775	3.4845	3.5062	Sapphire	Compressive	MOVPE	Reflectance	[32.222]
3.485	3.493	3.518	Sapphire	Compressive	MOCVD	Reflectance	[32.238]
3.491	3.499	3.528	Sapphire	Compressive	MOCVD	Reflectance	[32.223]
3.4903	3.4996	3.525	Sapphire	Compressive	MBE	Reflectance	[32.224]
3.484	3.490	3.512	Sapphire	Compressive	MOCVD	Contactless electro reflectance	[32.239]
3.479	3.486		Sapphire	Compressive	MOCVD	PL	[32.240]
3.488	3.496		Sapphire	Compressive	MOCVD	PL	[32.241]
3.4857	3.4921		Sapphire	Compressive	MOCVD	PL	[32.242]
3.483	3.489		Sapphire	Compressive	MBE	PL	[32.243]
3.476	3.489	3.511	ZnO	Compressive	RAMBE	Reflectance	[32.244]
3.480	3.493		ZnO	Compressive	RMBE	PL	[32.245]
3.470	3.474	3.491	6H—SiC	Tensile	MOCVD	Reflectance	[32.246]
3.470		3.486	6H—SiC	Tensile	MOVPE	PL	[32.247]
3.472			6H—SiC	Tensile	HVPE	PL	[32.248]
3.4771	3.4818	3.4988	GaN	Unstrained	MOVPE	Reflectance	[32.226]
3.4776	3.4827	3.5015	GaN	Unstrained	MOCVD	Reflectance	[32.225]
3.478	3.484	3.502	GaN	Unstrained	MOCVD	Reflectance	[32.229]
3.4772	3.4830	3.4998	Bulk GaN	Unstrained	Na—Ga melt	Reflectance	[32.249]
3.490	3.500	3.520	Freestanding	Unstrained	HVPE	Contactless electro reflectance	[32.250]

gap), and also a decrease in the A, B, and C splittings. As a summary, a comparison of free-exciton energies obtained from reflectance PL and other optical techniques for heteroepitaxial GaN grown by several techniques on various substrates, producing different degrees of strain, and unstrained homoepitaxy as well as bulk GaN is given in Table 32.5.

It should be pointed out that considerable variation exists between different reports in the literature [32.232, 238, 251]. The temperature dependence of the intrinsic-exciton energies is also dependent on the particular sample and local strain. Figure 32.24 presents a comparison between the temperature dependence of the exciton energies for three samples, presumably relaxed (bulk GaN), under compression (grown on sapphire) and under tension (grown on SiC). The temperature dependence of excitonic resonance (in the absence of localization) can be described by the Varshni empirical relation

$$E_g(T) = E_0(0) - \alpha T^2 / (\beta + T), \quad (32.15)$$

where $E_0(0)$ is the transition energy at 0 K, and α and β are the temperature coefficients. The free-exciton transitions are the dominating PL process at room temperature in GaN. This has been well established by PL spectral data over a wide range of temperatures for nominally undoped samples. However, the PL intensity at the position of the band gap at 300 K is considerably lower than the A exciton intensity [32.252].

Free-exciton transitions in wide-band-gap materials have a characteristic coupling to LO phonons [32.253]. As expected from the theory of LO-phonon coupling for exciton polaritons the first two replicas are strongest. The 3[−]LO and 4[−]LO replicas are also clearly observable. As predicted by theory, the characteristic temperature dependence of relative ratios of the intensities of the LO-phonon replicas is linearly proportional to the temperature and this behavior is confirmed for the A exciton for temperatures $T < 100$ K [32.248].

The binding energy of excitons determines the energy of the band gap of the material, but the strength of

the binding is also an important factor for the thermal stability of the excitons. *Chen et al.* [32.254] have calculated the binding energies as $\Delta E_A^b = \Delta E_B^b = 20$ meV for the A and B excitons and $\Delta E_C^b = 18$ meV for the C free exciton. By using the values of the effective electron and hole masses, *Chichibu et al.* [32.241] reported the calculated binding energy of the A exciton as 27 meV, which is close to the early estimation of 27.7 meV reported by *Mahler and Schroder* [32.255]. There have been very scattered experimental values reported for the free-exciton binding energies in GaN, as recently reviewed [32.256]. The scatter in the data is probably due to measurements on highly strained layers, and in some case misinterpretations of some features in the optical spectra. There is, in general, no feature at the band-edge position in the optical spectra, therefore, the exciton binding energy has to be obtained indirectly using the temperature dependence of the free-exciton transition in photoluminescence [32.240] or the position of the excited states if the positions of the excited states of the excitons are known [32.242, 243, 257]. The more recent data on both reflectance and photoluminescence (PL) reported for homoepitaxial samples appear convincing.

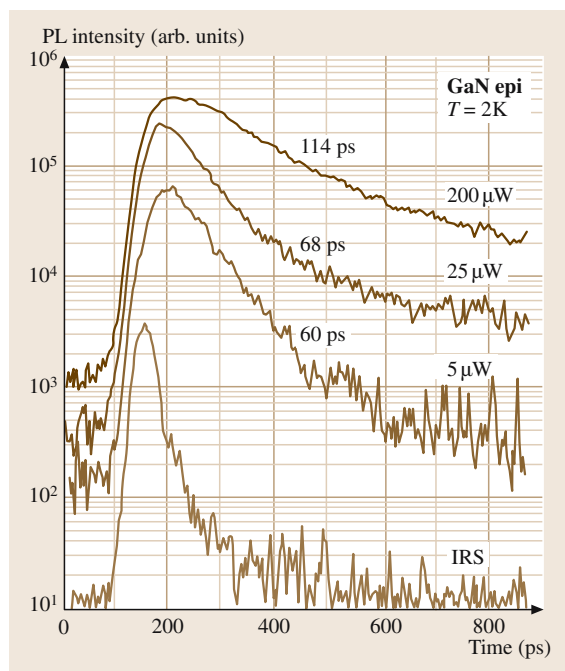


Fig. 32.25 Photoluminescence transients at 2 K for a GaN epilayer grown on sapphire measured at different excitation power intensity. IRS denotes the instrumental response in the experiment. [32.220]

A value of 25 meV is reported for the A exciton, and very similar values are reported for the B and C excitons [32.226, 258]. The $n = 2$ transitions are clearly resolved in these spectra, giving confidence that these results will be accurate for pure unstrained material, with a precision of about 1 meV.

The temporal behavior of excitons is of importance for emitters in that it provides a window on the dynamics of recombination processes. They should ideally be fast to be able to compete efficiently with nonradiative processes, such as multi-phonon emission and defect-related nonradiative centers. PL transient data represent an excellent tool to study both recombination rates of radiative and nonradiative processes in GaN. The temporal behavior at low excitation levels indicates processes, depending on the sample, that are as fast as about 35 ps for free excitons. Typical transients are shown in Fig. 32.25, which were obtained at different excitation intensities for a GaN epilayer grown on sapphire. Decay times at 2 K were found to vary between 60 ps and 115 ps, depending on the excitation intensity, indicating strong defect participation under these excitation intensities. The decay in luminescence intensity can more reasonably be described with a combination of a fast decay followed by a slower process with a decay time of about 300 ps at longer times. This slow process may be associated with weak localization of the free excitons, perhaps due to potential fluctuation induced by the inhomogeneous strain field.

Bound Excitons in GaN

The A, B, and C excitons discussed above represent intrinsic processes, as they do not involve pathways requiring extrinsic centers. In GaN, the neutral shallow donor-bound exciton (DBE), or I_2 , is often dominant because of the presence of donors, due to doped impurities and/or shallow donor-like defects. In samples containing acceptors, the acceptor-bound exciton (ABE), or I_1 , is observed. The recombination of bound excitons typically gives rise to sharp lines, with a photon energy characteristic for each defect. As in the free-exciton case, some fine structure is also expected in bound exciton lines, which are usually on the order of or below 1 meV for shallow donor- and acceptor-bound excitons. A characteristic phonon coupling, which can involve both lattice modes and defect-related vibrational modes, is also seen for each particular bound exciton spectrum. The photon energy region for DBE spectra is about 3.470–3.4733 eV at 2 K for strain-free GaN [32.218, 226]. Thick heteroepitaxial layers grown by HVPE give the best spectroscopic characteristic,

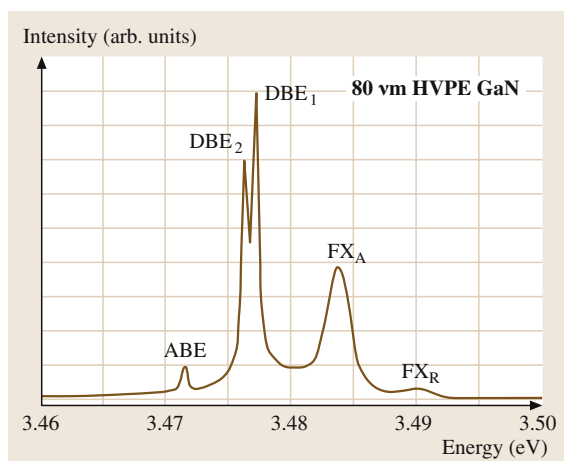


Fig. 32.26 Photoluminescence spectrum of a thick HVPE GaN layer showing two donor BE lines and an acceptor BE. The spectrum is upshifted by the compressive strain by about 6 meV compared with strain-free GaN. After [32.266]

with an optical line width of about 1 meV or less, as seen in Fig. 32.26. It is commonly argued that oxygen and silicon are the two most dominant residual shallow donors in GaN. In homoepitaxial GaN samples grown by MOVPE on a pretreated GaN single crystal, a narrower spectral line width down to 0.1 meV for the DBE peaked at 3.4709 eV has been reported [32.226]. In this case, in addition to unidentified sharp satellite lines on both sides of the main DBE peak, shallow DBE is the dominant line and has recently in the literature been assigned to substitutional Si donors [32.259]. Therefore, the high-energy line would then be O-related; O is known to be a typical contaminant in GaN grown by all techniques. Another high-quality MBE-grown GaN epitaxy on a GaN substrate shows at least three well-resolved peaks in the DBE region, at about 3.4709, 3.4718, and 3.4755 eV [32.260, 261]. In addition to the mostly accepted Si- and O-related shallow donor states, another donor is needed to identify the third line. It has been argued that these two levels (3.4709 and 3.4718 eV) could be attributed to exchange-split components of the same donor. Strain-splitting has also been considered in some cases. The most energetic one at 3.4755 eV is related to the B valance band involved in DBE. The localization energy for donor-bound excitons is about 6–7 meV. The corresponding binding energy of the dominating shallow donor electron has been estimated as about 35.5 meV from IR absorption data. An activation energy for electron conduction of

30–40 meV has been reported for O implanted in GaN by Pearton et al. [32.262]. Recent data from a combination of electrical and optical measurements on Si-doped GaN samples indicate that the Si donor has a binding energy of about 20 meV, while a deeper donor (35 meV) was also present, suggested to be O-related [32.151]. Monemar [32.263] estimated the binding energy of the donor electron from the two-electron satellites, which involves the radiative recombination of one electron with a hole, leaving the remaining neutral donor with a second electron in an excited $n = 2$ state. Assuming ideal effective-mass behavior, a binding energy of about

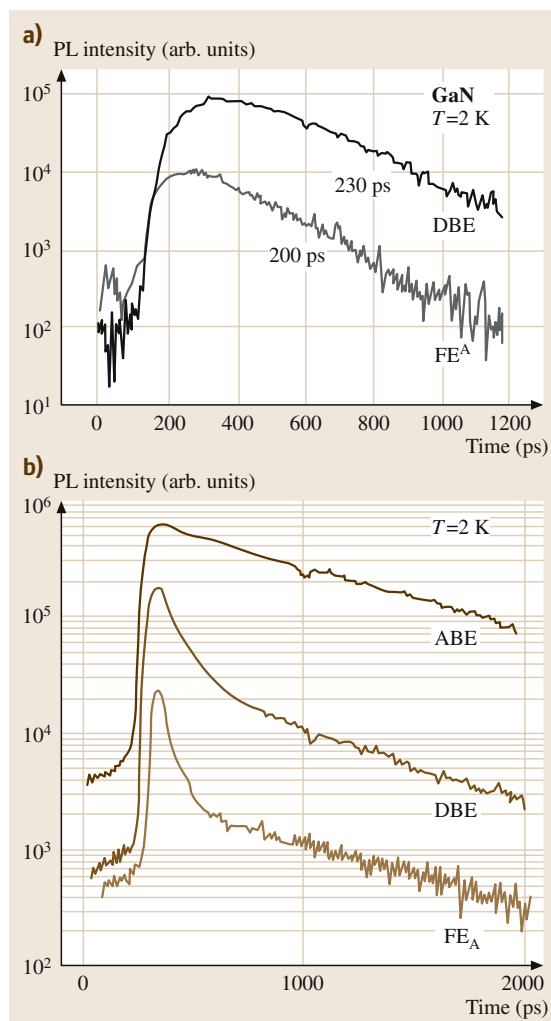


Fig. 32.27a,b Photoluminescence transients for a 400- μm -thick GaN layer (a) and for a thin strained layer (b) both grown on sapphire. After [32.266]

29 meV for the dominant donor in a 400- μm -thick HVPE GaN layer grown on sapphire has been estimated. By using the same technique for the MBE-grown GaN on a freestanding GaN template, Reshchikov et al. [32.264] determined the binding energies of two distinct shallow donors as 28.8 and 32.6 meV, which were attributed to Si and O, respectively. These combined data also provide strong evidence for a tentative identification of the 3.4709 eV DBE as being due to the neutral silicon donor. It is likely that the often-observed second DBE peak at 3.4718 eV in unstrained GaN is then related to the O donor, but this remains to be proven. In his review, Viswanath [32.265] listed commonly observed and reported peak positions and localization energies of DBE in GaN grown by various growth techniques on different substrates under compressive or tensile strain.

Another important property of DBEs is the recombination dynamics. The low-temperature decay rate for the DBE PL line gives the radiative lifetime of the DBE. Figure 32.27 shows photoluminescence transients measurements for a thick HVPE-grown GaN layer ($\approx 100\ \mu\text{m}$), with low dislocation density (in the $10^7\ \text{cm}^{-2}$ range), as well as for a thin strained layer grown on sapphire. In HVPE-grown thick layers, a low intensity decay time of about 200 ps is observed at 2 K, which is a typical radiative lifetime for DBEs related to shallow neutral donors. In thin strained heteroepitaxial layers with a higher defect density, a shorter lifetime is observed [32.267–269], indicating the effect of excitation transfer from the DBE to lower-energy states before the radiative recombination takes place. Similarly, homoepitaxial GaN layers also often show a fast DBE decay at low excitation density, due to excitation transfer to point defects [32.220]. A slow component, which is indicative of a slower radiative process overlapping the DBE transition, is also observed in the DBE decay dynamics [32.220, 257]. A recent observation on the DBE decay time for n-type GaN layers grown on SiC substrates indicated a clear trend versus the energy position of the DBE line, which in turn correlated with the strain in the layer. The biaxial strain in the GaN layer strongly affects the top valence band states, which in turn is reflected in the DBE wavefunction [32.258], affecting the oscillator strength and hence the observed radiative lifetime [32.227].

In the energy region below the principle DBE line there is usually a rich spectrum of bound excitons assumed to be acceptor-related. However, the situation for acceptor BEs (ABE or I_1) in GaN is somewhat less clear than for the donors. The most prominent neutral ABE is found at about 3.466 eV in strain-free GaN and

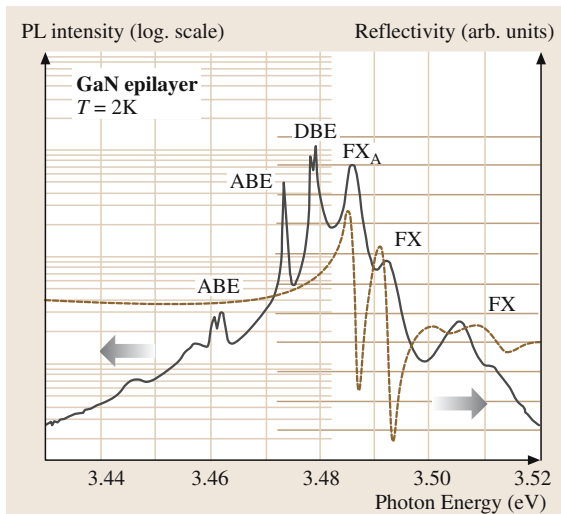


Fig. 32.28 Low-temperature PL spectrum (solid line) of a HVPE GaN layer shown on a semi-logarithmic scale. The reflectivity spectrum (dashed line) of the same sample taken at normal incidence geometry is plotted on a linear scale. After [32.266]

attributed to the exciton bound to a shallow acceptor, probably Mg_{Ga} (Fig. 32.28). Furthermore, this peak has been found to be dominant in slightly Mg-doped GaN samples [32.270–272]. An alternative interpretation of this BE line, which is related to charged donor-bound excitons, has recently been proposed by several groups [32.273–276]. The other main peak at about 3.455 eV has been attributed to Zn, but this is not yet justified. The broader peak at 3.44 eV is probably due to a low-energy acoustic-phonon wing connected to the main peak at 3.455 eV, which is very characteristic for ABEs [32.277, 278]. Also, phonon replicas are seen in the spectrum, indicating a rather strong LO-phonon coupling, much stronger than for the DBE. Since most experiments with doped samples up to now have been performed on heteroepitaxial material, there is a difficulty in determining the peaks, where the ABE peak position strongly depends on the strain. A binding energy of 19 meV was reported for Mg-doped GaN with an A exciton position at 3.499 eV and the Mg-related ABE at 3.480 eV at 2 K [32.279] indicating that ABEs, being deeper, may not follow a constant distance to the A free-exciton position (as in the case for the most shallow DBEs) in strained layers. In unstrained samples, like Mg-doped homoepitaxial GaN epilayers, the binding energy is found to be 11–12 meV [32.226, 280, 281], under the assumption that the 3.466 eV BE is Mg-related.

An additional deeper, probably acceptor-related, BE PL line in GaN is also observed at 3.461 eV (a doublet structure) [32.269]. However, due to the strain-induced shift in an upwards direction in the spectrum, this might be the same ABE as observed in some homoepitaxial layers at 3.455 eV with a line width of about 95 μeV [32.226], indicating a BE binding energy of about 21 meV. Although the identity of this acceptor has not been established, it is close to the position observed from the dominant ABE in Zn-doped samples, if strain shifts are considered [32.279]. Therefore, it might be due to residual Zn acceptors, present as contaminants in many samples, but this is not certain.

In contrast to the case of DBEs, the decay curves are usually clean exponential for the ABEs, as seen in Fig. 32.27b, and presumably reflect the radiative lifetimes of the BEs at the lowest temperatures, before thermalization occurs. The observed value of radiative lifetimes is about 0.7 ns for the shallowest ABE peaked at 3.466 eV, as compared to the much longer time, 3.6 ns, for the deeper acceptors with an ABE peak in the range 3.455–3.46 eV [32.269]. This corresponds to an oscillator strength of the order of 1, very similar to the shallow acceptor BEs in CdS [32.282]. To distinguish band-edge exciton features, the temperature-dependent luminescence experiment allows one to discriminate free versus donor-bound or acceptor-bound excitons. With increasing temperature, all peaks related to the bound excitons quench due to thermal delocalization, while the quenching of the free excitons is negligible up to 50 K. Excitons bound to the donor quench faster than excitons bound to the acceptor, which is consistent with their binding energies [32.264].

Donor–Acceptor Transitions in GaN

Due to compensation in semiconductors both ionized donors and acceptors could be present in the material. Nonequilibrium carriers generated by optical excitation can be trapped at the donor and acceptor sites, causing them to be neutral. While reaching equilibrium, they can relax their excess energies through radiative recombination of some electrons on the neutral donor sites with holes on the neutral acceptor sites, a process termed donor–acceptor pair (DAP) transition. DAP spectra are very common examples of radiative recombination in GaN. A no-phonon replica is observed at about 3.26 eV at low temperature, with well-resolved LO-phonon replicas towards the lower energy [32.283–285]. The temperature dependence of the spectrum reveals the evolution from a DAP pair spectrum at low temperature to a free-to-bound conduction-band-to-acceptor transition

peaked 3.273 eV at higher temperature (120 K) due to thermal ionization of the shallow donors ($E_D \approx 29$ meV) into the conduction band. At intermediate temperatures, both processes may be resolved. The acceptor binding energy has also been estimated as about 230 meV from this measurement. The identity of this acceptor is not clear, but it has been suggested that it is due to carbon acceptor substitution on nitrogen sites [32.285], while others claimed that it is simply related to Mg acceptors on Ga sites. PL of Mg-doped wurtzite GaN has been studied and showed the DAP transition at 3.26 eV at 4.2 K along with its LO-phonon replicas. At higher Mg concentrations, the PL is dominated by a deep-level broad band with its peak at 2.95 eV. The blue emission is clearly observed for Mg concentrations of 5×10^{19} – $2 \times 10^{20} \text{ cm}^{-3}$. In addition to the peak position for the identification of this transition, further confirmation is provided by a blue shift at a rate of about 2–3 meV per decade of intensity as the excitation intensity increased [32.286].

The behavior of radiative lifetimes and time-resolved spectra has been used as a further test for the DAP transition. It has been observed that carrier dynamics of 3.21 and 2.95 eV emissions in relatively heavily doped Mg-doped p-type GaN epilayers [32.287] exhibited nonexponential lifetimes on the sub-nanoseconds scale, comparable to that of a band-to-impurity transition in a highly n-type GaN epilayer involving a donor and the valence band [32.242, 288]. Smith et al. [32.287] observed the temporal behavior of the 3.21-eV emission band, which follows a power law with an exponent greater than 1.0 at longer delay times. All these indications lead one to conclude that the 3.21-eV line corresponds to the conduction-band-to-impurity transition involving shallow Mg impurities, while that at about 2.95 eV is attributed to the conduction-band-to-impurity transitions involving deep-level centers (or complexes).

Defect-Related Transitions in GaN: Yellow Luminescence (YL)

A broad PL band peaking at 2.2 ± 0.1 eV, so-called yellow emission, is almost systematically observed in undoped or n-GaN. As studied extensively by Ogino and Aoki [32.289] and more recently by others [32.290–292] there seems to be agreement that transitions from the conduction band or a shallow donor to a deep acceptor are responsible for this band. Another interpretation of the yellow band as being due to transitions between a deep donor and a shallow acceptor was also proposed to explain the results of magnetic-resonance experiments [32.293]. The exact position of this band

and its line width differ slightly in numerous publications. The nature of the deep donor has not been established, but several candidates such as the nitrogen vacancy [32.294], gallium vacancy or its complexes with a shallow donor [32.291, 292] or carbon [32.289, 295] have been suggested. The issue of whether the YL is related to a point defect or to a distribution of states in the gap is still an open question. *Shalish et al.* [32.296] invoked a broad distribution of acceptor-like surface states to account for the YL band. A redshift of the YL band with decreasing energy of the below-gap excitation may also indicate that the broadening of the YL band is due to emission from several closely spaced traps [32.297]. On the other hand, the temperature dependence of the bandwidth and the photoluminescence excitation spectrum have been quantitatively explained by using a configuration-coordinate model that attributes the YL band to a point defect with strong electron–phonon interaction [32.289, 298]. Theoretical calculations predict a low formation energy and the deep acceptor levels for the isolated Ga vacancy, possibly as a complex with a shallow donor impurity such as Si and O ($V_{Ga}Si_{Ga}$ and $V_{Ga}O_N$) [32.291, 292]. It has also been demonstrated that the formation of these defects is much more favorable at a threading-edge dislocation [32.299]. The intensity of the YL band has been found to correlate with dislocation-related defects [32.300].

The transient PL decay times of the YL has been investigated by different research groups. *Hofmann et al.* [32.301] and *Korotkov et al.* [32.302] reported rather long and nonexponential decay of the YL in the range 0.1–100 μ s at low temperatures, which was quantitatively described in the Thomas–Hopfield model [32.303] of the donor–acceptor pair (DAP) recombination. A very long-lived emission decay time of about 300 ms at 10 K has been observed by *Seitz et al.* [32.304]. In contrast, very fast decay times of about 1 ns at 2 K and 20 ps at room temperature have been reported and related to a strong contribution from free-to-bound transitions and DAP recombination by *Godlewski et al.* [32.305] and *Haag et al.* [32.306], respectively. *Reshchikov et al.* [32.307] performed time-dependent PL of YL in freestanding high-quality GaN templates grown by HVPE, which led them to conclude that, at temperatures below 40 K, the time decay of the 2.4 eV yellow peak is nonexponential and can be explained in the framework of the Thomas–Hopfield model for DAP-type recombination involving shallow donors. At elevated temperatures, the decay becomes exponential with two components, leading to the suggestion that

the transitions from the conduction band to two deep acceptors are involved.

32.6.2 Aluminium Nitride

The optical properties of AlN have been investigated in many forms, including powders, sintered ceramics, polycrystals and single-crystal samples. Since an AlN lattice has a very large affinity to oxygen dissolution, oxygen contamination is hard to eliminate in AlN, in which optical properties are influenced by oxygen-related defects. Some oxygen is dissolved in the AlN lattice while the remainder forms an oxide coating on the surface of each powder grain.

AlN doped with oxygen was found to emit a series of broad luminescence bands at near-ultraviolet frequencies at room temperature, no matter whether the sample was powdered, single crystal, or sintered ceramic. In an early study of the luminescence properties, *Pacesova and Jastrabik* [32.308] observed two broad emission lines centered near 3.0 and 4.2 eV, more than 0.5 eV wide, for samples contaminated with 1–6% oxygen under steady-state excitation. *Youngman and Harris* [32.309] and *Harris et al.* [32.310] found broad peaks centered at 2.7 and 3.8 eV in large single-crystal AlN with an oxygen content of 380 ppm. Oxygen-related luminescence spectra in AlN are very sensitive to sample preparation, particularly oxygen-impurity content. They observed an emission-peak shift in the PL spectra and a drastic increase in luminescence intensity below a critical oxygen content of 0.75%. Others reported different emission peaks for different forms of AlN samples with various oxygen contents. A microscopic model explaining the results was proposed by *Harris et al.* [32.310] and supported by the study of *Katsikini et al.* [32.311] and *Pastrnak et al.* [32.312]. Below the critical concentration, oxygen substitutes into nitrogen sites (O_N) with subsequent formation of Al atom vacancies (V_{Al}), while at higher oxygen content a new defect based on octahedrally coordinated Al forms.

Besides native oxygen defects or intentionally oxygen-doped AlN, other impurities that have been widely investigated are manganese and, more recently, rare-earth metals, such as erbium. In an early study, *Karel* and coworkers [32.313] reported a number of sharp emission peaks in the visible region from Mn-doped AlN. These peaks are interpreted as arising from phonon emission, in all likelihood due to localized Mn-ion vibration, associated with electronic transitions experienced by Mn^{4+} ions located at Al sites. This work was further extended by *Karel* and

Mares [32.314] and Archangelskii et al. [32.315], who observed a red band emission (600 nm) with Mn^{4+} and a green band (515 nm) with Mn^{2+} . Recently, particular attention has been paid to Er-doped AlN due to the observed strong photoluminescence at 1.54 μm , which is important for optical-fiber communication systems. In rare-earth atoms, the luminescence transition comes from the core electrons falling energetically from excited states to the ground states. For example, Er-doped AlN gives a luminescence due to transitions between the weak crystal-field-split levels of $\text{Er}^{3+}I_{13/2}$ and $^4I_{15/2}$ multiplets [32.316–318]. Pearton et al. [32.319] reported luminescence enhancement in AlN(Er) samples treated in a hydrogen plasma by passivating the defects in AlN. This increase shows no quenching on heating the sample up to 300 °C. Due to its wide band gap the 1.54- μm luminescence is more stable in AlN than in GaN.

In recent work, band-edge emission of high-quality AlN grown by MOCVD on sapphire substrate has been investigated by Li et al. [32.320]. Band-edge emission lines at 5.960 and 6.033 eV have been observed at room temperature and 10 K, respectively, as shown in Fig. 32.29. The peak integrated emission intensity of the deep-impurity-related emission centered at around 3.2 and 4.2 eV is only about 1% and 3%, respectively, of that of the band-edge transition at room temperature. The PL emission properties of AlN have been compared with those of GaN and it was shown that the optical quality as well as the quantum efficiency of AlN epilayers is as good as that of GaN. The same group [32.322] has also studied deep UV picosecond time-resolved photoluminescence spectroscopy to investigate the observed optical transition in steady-state PL measurements. Two PL emission lines at 6.015 and 6.033 eV were attributed to donor-bound exciton and

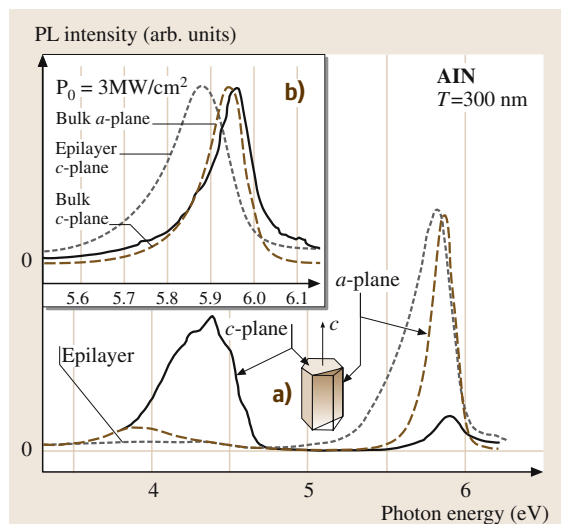


Fig. 32.30 (a) PL spectra of single-crystal AlN for A- and C-planes and AlN epilayers at room temperature under an excitation power density of 1 MW/cm². Inset (b) shows wurtzite-type crystal c-axes, C-plane, and A-plane. Inset shows room-temperature near-band emission spectra of various AlN samples under very high excitation. After [32.321]

free-exciton transitions, respectively, from which the binding energy of the donor-bound excitons in AlN epilayers was determined to be around 16 meV.

Kuokstis et al. [32.321] reported similar results by employing temperature-dependent PL measurements for A-plane and C-plane bulk single crystals and MOCVD-grown epitaxial layers on sapphire. The spectra consisted of a long wave band probably from an oxygen-related deep-level transition as discussed above and an intense

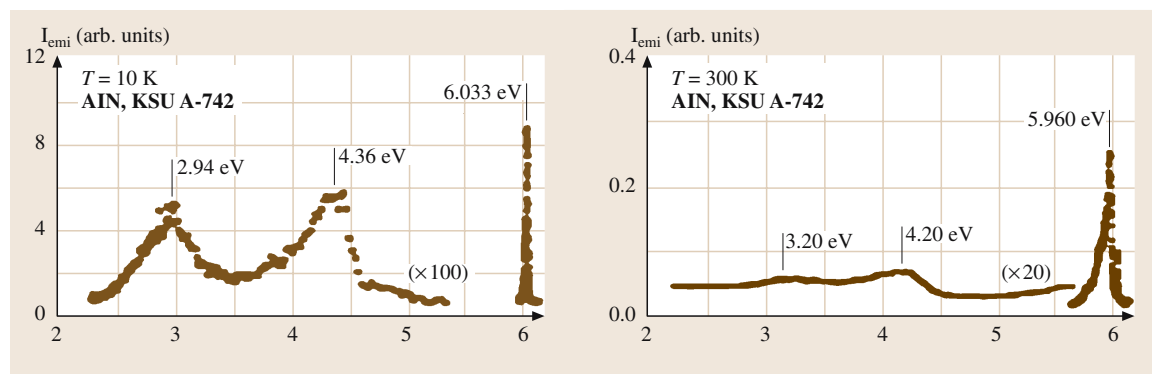


Fig. 32.29a,b PL spectra of AlN epilayers measured (a) at 10 K and (b) at 300 K. The spectra cover a broad spectral range from 2.2 to 6.2 eV. After [32.320]

near-band emission, as seen in Fig. 32.30. The near-band emissions were attributed to processes involving free excitons at 6.065 eV and their LO-phonon (≈ 5.066 eV) replicas at high temperatures and the bound excitons at 6.034 eV at low temperatures. These results, together with other well-known physical properties of AlN, may considerably expand future prospects for the application of group III nitride materials.

Time-resolved PL measurements revealed that the recombination lifetimes were around 80 ps for bound excitons and 50 ps for free excitons [32.322]. A free-exciton binding energy of 80 meV was also deduced from the temperature dependencies of both the free-exciton radiative decay time and emission intensity.

On the other hand, optical reflectance (OR) measurements [32.323] exhibited distinct reflectance anomalies at photon energies just above the multiple-internal-reflection fringes, and the spectral line shape was fitted considering A, B, and C exciton transitions. The fit gave values for the A and B excitons at 0 K of 6.211 and 6.266 eV, giving a crystal-field splitting (Δ_{cr}) of approximately 55 meV. The AlN film exhibited an excitonic emission even at 300 K, which is due to the small Bohr radius of the excitons and the large longitudinal optical-phonon energies.

Cathodoluminescence and electroluminescence emission properties of AlN have also been investigated by several groups. Rutz [32.177] reported a broad near-UV band in the electroluminescence spectra, extending from 215 nm into the blue end of the visible region. Several peaks or humps between 2.71 and 3.53 eV, which were attributed to nitrogen vacancies, interstitial Al impurities, and oxygen impurities or defects, were observed in cathodoluminescence measurements for epitaxial AlN films grown on C-plane sapphire [32.324]. In the investigation of Youngman and Harris, 4.64-eV photons were used specifically to explore the below-band transitions such as the peak near 4 eV. Their investigation showed a blue shift with increased O concentration in the peak in question. Hossian et al. [32.325] and Tang et al. [32.326] measured the cathodoluminescence at 300, 77, and 4.2 K for undoped AlN thin films grown on sapphire and SiC by LP-MOCVD (low pressure). From these samples, a strong luminescence peak surrounded by two weaker peaks in the near band-edge region, near 6 eV, was observed. For AlN on sapphire, this near-band-edge transition can be further resolved into three peaks at 6.11, 5.92, and 5.82 eV. They believe that these two peaks are due to excitonic transition because of their exciton-like temperature behavior, i. e. the energy position of these peaks increases and the

line width becomes narrower as the temperature is decreased. Recently, Shishkin et al. [32.327] has also observed a sharp, strong near-band-edge peak at 5.99 eV, which was tentatively assigned to the optical recombination of a donor-bound exciton, accompanied by weak one- and two-longitudinal optical-phonon replicas in the cathodoluminescence spectrum of a 1.25- μm -thick RMBE-grown (reactive molecular beam epitaxy) AlN film on C-plane sapphire.

Not much data are available on the temperature dependence of the optical band gap of AlN. Guo and Yoshida [32.328] investigated the variation of the band gap with temperature in AlN crystalline films. The band-gap energy increases linearly from room temperature down to about 150 K with a temperature coefficient of 0.55 meV/K. Another work was carried out by Kuokstis et al. [32.321] and similar temperature behavior was observed for single-crystal C-plane AlN with a temperature coefficient of 0.53 meV/K between 150 and 300 K (Fig. 32.31). These two observations are close to each other and also close to the value reported by Tang et al. [32.326] (0.51 meV/K). However, below 150 K, in all measurements, a new structure starts to form, where the band gap does not change linearly with temperature and has a very small temperature coefficient, which is unusual for semiconductors.

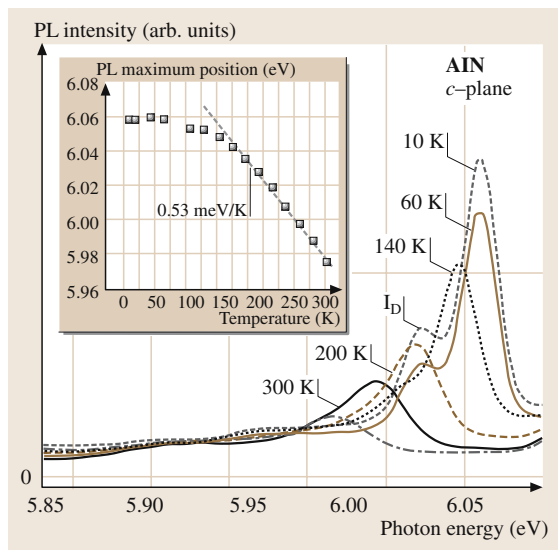


Fig. 32.31 Single-crystal AlN C-plane PL spectra at different temperatures. B denotes the free-exciton line, whereas I_D marks the bound-exciton line. The inset shows the dependence of the exciton line position on temperature. After [32.321]

In the case of absorption, the absorption coefficient in the near band edge was measured by using straightforward optical transmission and reflection experiments. In early studies, *Yim et al.* [32.174] characterized AlN by optical absorption and determined the room-temperature band gap to be direct with a value of 6.2 eV. *Perry and Rutz* [32.329] performed temperature-dependent optical absorption on a $1 \times 1 \text{ cm}^2$ epitaxial single-crystal AlN sample, measuring a band gap of 6.28 eV at 5 K compared to their room-temperature value of 6.2 eV, resulting from a straight-line fit. Several groups have reported comparable values whereas others have produced questionable values considerably below 6.2 eV, probably due to oxygen contamination or nonstoichiometry. In addition to the band-edge absorption, a much lower energy absorption peak at 2.86 eV (although some variation in the peak position has been recorded from 2.8 to 2.9 eV) is likely to be due to nitrogen vacancies or nonstoichiometry, as proposed by *Cox et al.* [32.173]. *Yim and Paff* [32.101] also observed a broad emission-spectrum range of 2–3 eV with a peak at about 2.8 eV. This peak does not correlate with the presence of oxygen. The oxygen absorption region lies between 3.5 eV and 5.2 eV, as originally found by *Pastrnak and Roskovicova* [32.330]. The exact position of this particular peak appears to change with the oxygen content from 4.3 eV at low oxygen levels to 4.8 eV at high oxygen levels.

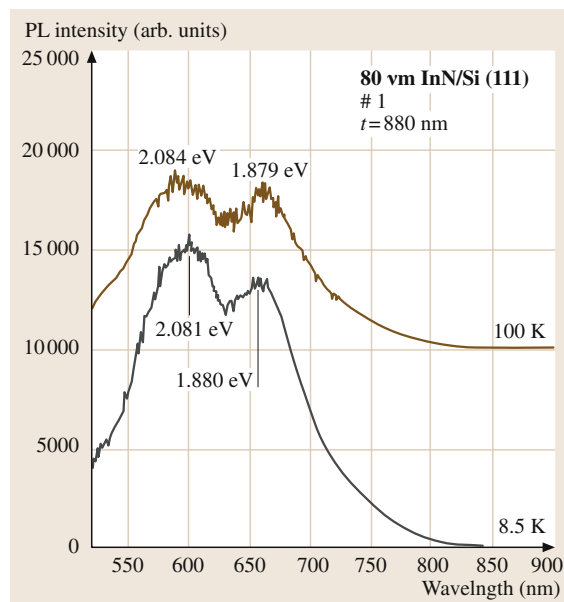


Fig. 32.32 PL spectra at 8.5 and 100 K of InN films grown on Si substrate. After [32.331]

Measurements of the refractive index of AlN have been carried out in amorphous, polycrystalline, and single-crystal epitaxial thin films [32.332–335]. The values of the refractive index are in the range $n = 1.9$ – 2.34 with a generally accepted value of $n = 2.15 \pm 0.05$ at visible wavelengths. These values are found to increase with increasing structural order, varying between 1.8 to 1.9 for amorphous films, 1.9 to 2.1 for polycrystalline films, and 2.1 to 2.2 for single-crystal epitaxial films. The spectral dependence and the polarization dependence of the index of refraction have been measured and showed a near-constant refractive index in the wavelength range 400–600 nm. The reported refractive index generally agrees with low- and high-frequency dielectric constants determined from infrared measurements. In the long-wavelength range, the low-frequency dielectric constant of AlN (ϵ_0) lies in the range 8.5–9.1 [32.336,337], and most of the values fall within the range 8.5 ± 0.2 , while the high-frequency dielectric constant (ϵ_∞) is in the range 4.6–4.8 [32.98,338].

32.6.3 Indium Nitride

Among the group III nitrides, InN is not very well investigated. The band gap has been frequently measured by transmission experiments. In early studies, due to the large background electron concentrations of the samples and their low crystalline quality, the excitonic features, neither in the absorption spectra nor in the band-edge photoluminescence spectra, had been observed. However, a number of groups have performed optical measurements on InN [32.339–342]. Early experimental studies suggested that the band gap of wurtzite InN ranges from 1.7–2.07 eV at room temperature. Table 32.1 lists some experimental determinations of the InN band gap. They were estimated from the absorption spectra, which can lead to an overestimation of the band gap if the sample quality is poor or if the sample is highly doped. Despite this, the value of 1.89 eV for the InN band gap obtained by *Tansley and Foley* [32.343] is widely cited in the literature. They explained this experimental scatter by considering residual electron concentration (which changes the band gap of the semiconductor, as discussed before) and attributed this variation to the combined effect of band-tailing and band-filling. The value of 1.89 eV is often used as an end-point value to interpret experimentally measured composition dependence of the band gap of InN alloys [32.344]. A few studies of the inter-band optical absorption performed on InN thin films deposited by sputtering techniques [32.343] and

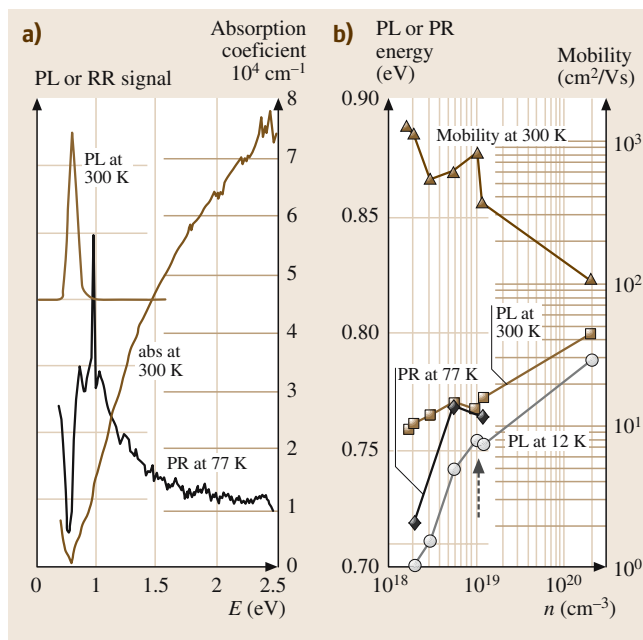


Fig. 32.33 (a) Optical absorption (300 K), PL (300 K), and PR (77 K) spectra of a typical InN sample. This sample is undoped with a room-temperature electron concentration of $5.48 \times 10^{18} \text{ cm}^{-3}$. The spike on the PR spectrum at 0.97 eV is an artifact due to the light source used in the PR measurement. (b) Room-temperature mobility, PL peak energy (300 and 12 K), and the critical energy determined by PR (77 K) as a function of free-electron concentration. After [32.349]

MOCVD [32.328] were found consistent with a fundamental energy gap of about 2 eV. Recent progress of growth techniques using molecular-beam epitaxy has led to improved InN samples, which show photoluminescence as well as a clear absorption edge. Yodo et al. [32.331] observed weak photoluminescence peaks with energies in the range 1.81–2.16 eV on InN grown on Si substrates (Fig. 32.32). However, in one such case, an emission centered at 1.86 eV at a temperature below 20 K was seen, while the reflectance measurement shows a strong plasma reflection at 0.7 eV. Another branch of studies show that in improved InN films strong photoluminescence transitions at energies around 1 eV appear [32.345–348]. These new measurements have challenged the previous widely accepted band-gap value and suggest that the actual fundamental band gap of InN is much smaller, between 0.7 and 1.1 eV. The real band gap of InN is still under investigation and undetermined.

Figure 32.33 shows the room-temperature electron mobility, the peak energy of PL, and the transition energy

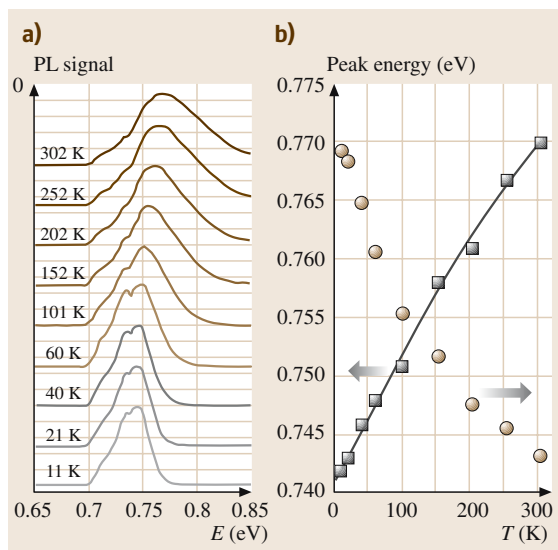


Fig. 32.34 (a) PL spectra as a function of temperature for the sample shown in Fig. 32.20a. PL spectra are normalized to a constant peak height. (b) PL peak energy and PL integrated intensity (log scale) as a function of temperature. The line through the peak energy data is a guide to the eye. After [32.349]

determined by PR as functions of electron concentration. The transition energies increase with increasing free-electron concentration, indicating that the transitions from higher-energy occupied states in the conduction band contribute significantly to the PL spectrum. The absorption coefficient increases gradually with increasing photon energy, and at the photon energy of 1 eV it reaches a value of more than 10^4 cm^{-1} . This high value is consistent with inter-band absorption in semiconductors. Moreover, the integrated PL intensity increased linearly with excitation intensity over three orders of magnitude, lending more credence to the concept that the observed nonsaturable peak is related to the fundamental inter-band transitions. The free-electron concentration in this sample was measured by the Hall effect to be $5.48 \times 10^{18} \text{ cm}^{-3}$, which also shows intense room-temperature luminescence at energies close to the optical absorption edge. Additionally, the 77-K photo reflectance (PR) spectrum exhibits a transition feature at 0.8 eV with a shape characteristic for direct-gap inter-band transitions. Consistent with the absorption data, no discernible change in the PR signal near 1.9 eV is seen. The simultaneous observations of the absorption edge, PL, and PR features at nearly the same energy indicate that this energy position of $\approx 0.78 \text{ eV}$ led Wu

et al. [32.349] to argue that this is the fundamental band gap of InN. This value is very close to the fundamental gap for InN reported by Davydov et al. [32.345].

A few groups [32.350, 351] have reported the temperature dependence of the band gap of InN, and indicated a band-gap temperature coefficient of $1.3\text{--}1.8 \times 10^{-4}$ eV/K at room temperature by using $E_g(300\text{ K}) = 1.89$ eV. The temperature dependence of PL spectra of recently investigated InN samples indicated a lower band-gap energy of 0.7 eV, as shown in Fig. 32.34. There is a small blue shift (nearly linear at 0.1 meV/K) observed in the peak energy position of the temperature-dependent PL. In addition to that, the integrated intensity of the PL decreases by ≈ 20 times as the temperature is increased from 11 K to room temperature. The data in Fig. 32.34a also show a considerable increase in the line width of the PL spectra. The full width at half maximum increases from 35 to 70 meV when the temperature increases from 11 K to room temperature.

32.7 Properties of Nitride Alloys

Many important GaN-based devices involve heterostructures to achieve improved device performance. Ternary alloys of wurtzite polytypes of GaN, AlN, and InN have been obtained in continuous alloy systems whose direct band gap ranges from ≈ 0.8 eV for InN to 6.1 eV for AlN. Many of these properties, such as the energy band gap, effective masses of the electrons and holes, and the dielectric constant, are dependent on the alloy composition. AlN and GaN are reasonably well lattice-matched (3.9%) and, owing to the large band-gap difference between GaN and AlN, for many devices only small amounts of AlN are needed in the GaN lattice to provide sufficient carrier and optical field confinement. The compositional dependence of the lattice constant, the direct energy gap, electrical and cathodoluminescence (CL) properties of the AlGaIn alloys were measured [32.358]. In general, the compositional dependence of the optical band gap of ternary alloys can be predicted by the following empirical expression;

$$E_g(x) = xE_g(\text{AlN}) + (1-x)E_g(\text{GaN}) - x(1-x)b, \quad (32.16)$$

where $E_g(\text{AlN})$, $E_g(\text{GaN})$ are the optical band gap of AlN and GaN, respectively, and x and b are the AlN molar fraction and bowing parameter, respectively. In order to determine this relation precise characterization of both the band gap and alloy composition is important. Wide dispersion in the bowing parameters ranging from

Therefore, they concluded that there is no significant shift of the PL spectra, as the temperature-induced line broadening can easily account for the observed small upward shift of the PL line maximum.

In the transparency region, the refractive index was measured using the interference spectrum observed in reflectivity or absorption [32.352, 353]. In the visible and UV region ($E > 2$ eV) optical constants were determined from reflectivity or spectroscopic ellipsometry [32.354, 355]. Although the most frequently quoted value for the static dielectric constant is $\epsilon_0 = 15.3$, the values reported for the high-frequency dielectric constant are scattered between 6.6 and 8.4, as well as the refractive index measured in the vicinity of the band gap. Tyagai et al. [32.356] were able to estimate an effective mass of $m^*_e = 0.11 m_0$ and an index of refraction $n = 3.05 \pm 0.05$ in the vicinity of the band gap. The long-wavelength limit of the refractive index was reported to be 2.88 ± 0.15 .

-0.8 eV (upward bowing) to $+2.6$ eV (downward bowing) has been reported [32.359–364]. Yun et al. [32.357]

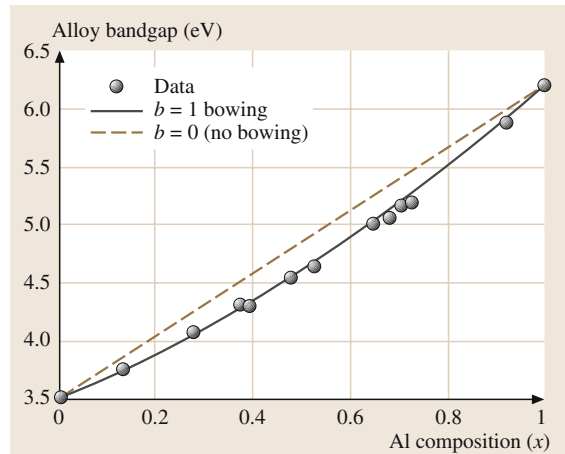


Fig. 32.35 Experimental data for the energy band gap of AlGaIn ($0 \leq x \leq 1$) plotted as a function of Al composition (solid circle), and the least-squares fit (solid line), giving a bowing parameter of $b = 1.0$ eV. The dashed line shows the case of zero bowing. As the quality of the near-50:50 alloy layers get better, giving rise to sharper X-ray and PR data, smaller bowing parameters may ensue. Already, bowing parameters as low as 0.7 have been reported. After [32.357]

revisited the bowing parameter using X-ray and analytical techniques, such as secondary-ion mass spectroscopy (SIMS) and Rutherford back-scattering (RBS), for composition determination, reflectance and absorption for band-gap determination. The results of this study are depicted in Fig. 32.35 in the form of AlGa_xN band gap versus composition along with a least-square fit to the data, solid circles, yields a bowing parameter of $b = 1.0$ eV for the entire range of alloy compositions. Widening the X-ray diffraction peaks for alloy composition around the midway point has been attributed to be a most likely source of error in determining the bowing parameter. On the other hand, the validity of the characterization techniques used in determination the optical properties of AlGa_xN alloys is deeply affected by the material crystalline quality and purity.

As far as the electrical and doping issues are concerned, Hall measurement for n-Al_{0.09}Ga_{0.91}N demonstrated a carrier concentration of $5 \times 10^{18} \text{ cm}^{-3}$ and a mobility of $35 \text{ cm}^2/\text{Vs}$ at room temperature [32.365]. Other Hall measurements [32.366] on Mg-doped p-Al_{0.08}Ga_{0.92}N grown by OMVPE (organometallic vapor phase epitaxy) addressed the temperature dependence of the mobility. They indicate that the hole mobility decreases with increasing temperature, reaching a value of about $9 \text{ cm}^2/\text{Vs}$ for a doping density of $1.48 \times 10^{19} \text{ cm}^{-3}$. This low mobility is ascribed to a high carrier concentration and the inter-grain scattering present in the samples. While the lattice constant was studied, it was observed to be almost linearly dependent on the AlN mole fraction in AlGa_xN.

Until recently the resistivity of unintentionally doped AlGa_xN was believed to increase so rapidly with increasing AlN mole fraction that AlGa_xN became almost insulating for AlN mole fractions exceeding 20%. As the AlN mole fraction increased from 0 to 30%, the n-type carrier concentration dropped from 10^{20} to 10^{17} cm^{-3} , and the mobility increased from 10 to $30 \text{ cm}^2/\text{Vs}$. An increase in the native-defect ionization energies with increasing AlN may possibly be responsible for this variation. The respond of the dopant atoms such as Si and Mg to the variation of the AlN mole fraction in AlGa_xN has not been well understood yet. It was suggested that dopant atom moves deeper into the forbidden energy band gap as the AlN mole fraction increases. For example, Hall-effect measurements show that the activation energy of Si donor increases linearly from 0.02 eV in GaN to 0.32 eV in AlN [32.367]. However, devices such as lasers, which depend critically on the overall device series resistance and require low-resistivity p-type material, will

probably be restricted by the ability to dope high-mole-fraction AlGa_xN. Fortunately, the emergence of InGa_xN coupled with the fact that good optical-field confinement can be obtained with low-AlN-mole-fraction AlGa_xN mitigate this problem enormously, and the potential is very bright for laser development in this material system.

The ternary InGa_xN is used mostly for quantum wells, strained to some extent depending on the level of phase segregation, etc., in the active regions of LEDs and lasers, which can emit in the violet or blue wavelength range. Needless to say, high-efficiency blue and green LEDs utilizing InGa_xN active layers are commercially available. However, added complexities such as phase separation and other inhomogeneities due to the large disparity between Ga and In make the determination of the band gap of InGa_xN versus composition a very difficult task, not to mention the controversy regarding the band gap of InN. The compositional dependence of the InGa_xN band gap is a crucial parameter in the design of any heterostructure utilizing this material. Similarly to the case of AlGa_xN, the energy band gap of In_xGa_{1-x}N over $0 < x < 1$ can be expressed by (32.16) using the band gap of InN instead of AlN band gap. When a bandgap of ≈ 1.9 eV for InN is assumed as the end-point value for InN in regard to the InGa_xN ternary, large and/or more than one bowing parameters are required to fit the compositional dependence of the band-gap energy. An earlier investigation of InGa_xN bowing parameter for alloys with small concentrations of InN [32.368] led to a bowing parameter of 1.0, which is in disagreement with 3.2 reported by *Amano* et al. [32.369], who also took into consideration strain and piezoelectric fields and arrived at a value of 3.2. A bowing parameter of 2.5 eV was obtained from optical absorption measurements and a value of 4.4 eV was obtained from the position of the emission peak [32.370]. *Nagatomo* et al. [32.371] noted that the In_xGa_{1-x}N lattice constant varies linearly with the In mole fraction up to at least $x = 0.42$, but it violates Vegard's law for $x > 0.42$, which may be caused by erroneous determination of the composition and is very relevant the problem at hand. Recent observations indicated that these alloys show strong infrared PL signal as expected from an InN band gap of ≈ 0.8 eV, extending the emission spectrum of the In_{1-x}Ga_xN system to near infrared. [32.347] revisited the dependence of the InGa_xN band gap on composition by considering ≈ 0.8 eV for the band gap of the end binary InN. Figure 32.36 shows the compositional dependence of the band gap of InGa_xN, determined by photomodu-

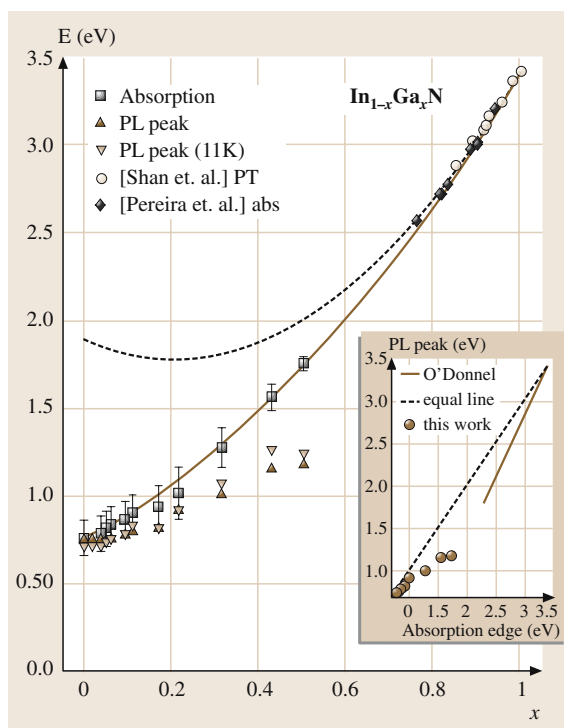


Fig. 32.36 PL peak energy and band gap determined by optical absorption as a function of composition, including previously reported data for the Ga-rich end of the ternary which is not as affected by the large InN band gap previously used. The *solid curve* shows the fit to the band gap energies (abs and PT) using a bowing parameter $b = 1.43$ eV. The *dashed curve* is the fit to the band-gap energies on the Ga-rich side assuming a band gap of 1.9 eV for InN. *Inset*: PL peak energy plotted against absorption edge energy. The *solid line* is a least-squares fit to experimental data on the Ga-rich side. The *dashed straight line* shows the relation when the Stokes shift is zero. After [32.245] and courtesy of Wlodek Walukiewicz

lated transmission [32.372] optical absorption [32.373] measurements, as a function of GaN fraction. The compositional dependence of the band gap in the entire compositional range can also be fit by a bowing parameter of $b = 1.43$ eV. Also shown in the figure with a dashed line is the fit to the empirical expression using an energy of 1.9 eV for InN and bowing parameter of 2.63 eV to demonstrate that it does represent the Ga-rich side of the compositions well. However, the bowing parameter that is good for the entire compositional range is 1.43 eV, utilizing 0.77 eV for the band gap of InN.

There are isolated reports on electrical properties of the InGaN alloy. Nagatoma et al. [32.371] reported high-resistivity InGaN films with In compositions as high as 0.42. Background electron concentration as low as 10^{16} cm^{-3} has been reported for MOVPE-grown InGaN films. Conductivity control of both n-type and p-type InGaN was reported by Akasaki and Amano [32.374]. The growth and mobility of p-InGaN was also discussed by Yamasaki et al. [32.375]. Yoshimoto et al. [32.376] studied the effect of growth conditions on the carrier concentration and transport properties of $\text{In}_x\text{Ga}_{1-x}\text{N}$. They observed that increasing the deposition temperature of $\text{In}_x\text{Ga}_{1-x}\text{N}$ grown both on sapphire and ZnO with $x \approx 0.2$ results in a decrease in carrier concentration and increase in electron mobility. Nakamura and Mukai [32.150] discovered that the film quality of $\text{In}_x\text{Ga}_{1-x}\text{N}$ grown on high-quality GaN films could be significantly improved. Thus, it may be concluded that the major challenge for obtaining high-mobility InGaN is to find a compromise value for the growth temperature, since InN is unstable at typical GaN deposition temperatures. This growth temperature would undoubtedly be a function of the dopant atoms, as well as the method (MBE, OMVPE, etc.) used for the growth. This is evident from a study by Nakamura and coworkers, who have since expanded the study of InGaN employing Si [32.377] and Cd [32.378].

$\text{In}_{1-x}\text{Al}_x\text{N}$ is an important compound that can provide a lattice-matched barrier to GaN, low-fraction AlGaIn and InGaIn, and consequently, lattice-matched AlInN/AlGaIn or AlInN/InGaIn heterostructures would result. The growth and electrical properties of this semiconductor have not yet been extensively studied, as the growth of this ternary is also challenging because of the different thermal stability, lattice constant, and cohesive energy of AlN and InN. Kim et al. [32.379] deposited thin AlInN films and observed an increase of In content in AlInN of up to 8% by lowering the substrate temperature to 600 °C. Radio-frequency (RF) sputtering was employed to grow InAlN alloy by Starosta [32.380], and later Kubota et al. [32.342]. Kistenmacher et al. [32.381], on the other hand, used RF-magnetron sputtering (RF-MS) from a composite metal target to grow InAlN at 300 °C. It was observed that the energy band gap of this alloy varies between 2.0 eV and 6.20 eV for x between 0 and 1 [32.342]. Optical properties of 1- μm -thick $\text{Al}_{1-x}\text{In}_x\text{N}$ layers for x values up to the range of 0.19–0.44 have been investigated by absorption and photoluminescence [32.382]. Figure 32.37 shows results of the photoluminescence spectra from $\text{Al}_{1-x}\text{In}_x\text{N}$ layers. From the absorption spectra the band

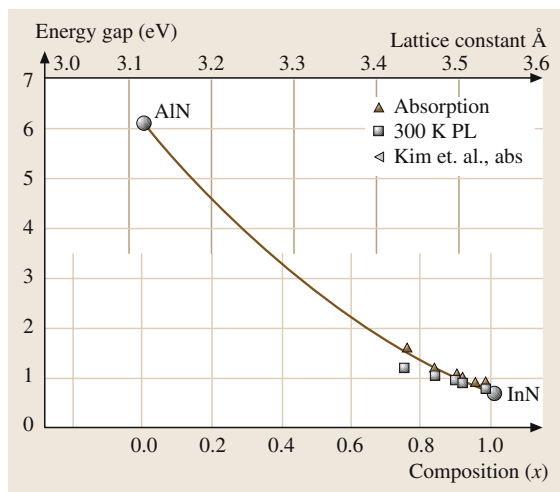


Fig. 32.37 Band gap of $\text{In}_x\text{Al}_{1-x}\text{N}$ films for various In concentrations. Courtesy of Wlodek Walukiewicz

gaps are found to shift gradually with varying In content. Tailing of the band gap is also found to occur, probably due to fluctuation of the In content and grain boundaries. The PL peak-energy position shifts towards the lower-energy region with increasing In content as expected [32.383]. Considering that PL spectral line widths are much larger than expected from a completely random alloy, mechanisms other than alloy broadening such as microscopic phase separation might be invoked to account for such broadening. The small amount which the PL peak shifts (the same is true for the absorption edge) with mole fraction, x , might reflect the immiscibility of AlN in InN (or vice versa).

Transport data for InAlN are extremely scarce. The carrier concentration and the mobility of $\text{In}_{1-x}\text{Al}_x\text{N}$ for $x = 0.04$ were reported as $2 \times 10^{20} \text{ cm}^{-3}$ and $35 \text{ cm}^2/\text{Vs}$, respectively, and $8 \times 10^{19} \text{ cm}^{-3}$ and $2 \text{ cm}^2/\text{Vs}$ for $x = 0.25$, respectively [32.384]. The mobility was found to decrease substantially with an increase in the Al mole fraction and increase with an

increase in the In mole fraction, in close analogy with the parent compounds AlN and InN.

By alloying InN with GaN and AlN, the band gap of the resulting alloy(s) can be increased from 1.9 eV ($\approx 0.8 \text{ eV}$ if we use the recently determined InN band gap) to a value of 6.1 eV, which is critical for making high-efficiency visible-light sources and detectors. In addition, the band gap of this quaternary can be changed while keeping the lattice constant matched to GaN [32.385, 386]. In quaternary alloys of nitrides the N atoms constitute an anion sublattice while the three group III elements (In, Ga, Al) constitute the cation sublattice. Use of this quaternary material allows almost independent control of the band offset in AlInGaN-based heterostructures. However, among other difficulties brought about by the four-component system, the optimal growth temperature is important to optimize and control, as aluminium-based compounds generally require higher growth temperatures, while In-based compounds require lower temperatures. Higher temperatures are also desirable to reduce the incorporation of O into the growing film as oxides of Ga and In desorb from the surface. The growth temperature will therefore govern the limits of In and Al incorporation into the AlGaInN quaternary alloy [32.385]. Quaternary alloys, $(\text{Ga}_{1-x}\text{Al}_x)\text{In}_{1-y}\text{N}$, are expected to exist in the entire composition range $0 < x < 1$ and $0 < y < 1$. Unfortunately, as is the case for the InGaIn alloy, the incorporation of indium into these quaternary alloys is not easy. Ryu et al. [32.387] reported on optical emission in this quaternary system and AlInGaIn multiple quantum wells grown by pulsed metalorganic chemical-vapor deposition. A strong blue shift with excitation intensity was observed in both the quaternary layers and quantum wells, which was attributed to localization. This would imply that the structures are of inhomogeneous nature and/or presence of band-tail states indicative of early stages of material development and/or serious technological problems involved.

32.8 Summary and Conclusions

AlN exhibits many useful mechanical and electronic properties. For example, hardness, high thermal conductivity, resistance to high temperature and caustic chemicals combined with, in noncrystalline form, a reasonable thermal match to Si and GaAs, make AlN an attractive material for electronic packaging applications.

The wide band gap is also the reason for AlN to be touted as an insulating material in semiconductor device applications. Piezoelectric properties make AlN suitable for surface-acoustic-wave devices and sensor applications. However, the majority of interest in this semiconductor in the context of electronic and optoelectronic devices

stems from its ability to form alloys with GaN, producing AlGaIn and allowing the fabrication of electronic and optical devices based on AlGaIn/GaN and AlGaIn/InGaIn, the latter of which can be active from green wavelengths well into the ultraviolet. AlN also forms a crucial component of the nitride-based AlInGaIn quaternary which allows tuning of the band gap independently of the composition over a reasonably wide range of band gaps. This way, lattice-matched conditions to the underlying epitaxial structure can be maintained while being able to adjust the band gap.

AlN is not a particularly easy material to investigate because of the high reactivity of aluminium with oxygen in the growth vessel. Early measurements indicated that oxygen-contaminated material can lead to errors in the energy band gap and, depending on the extent of contamination, in the lattice constant. Only recently has the attainment of contamination-free deposition environments coupled with advanced procedures allowed researchers to grow higher-quality AlN consistently. Consequently, many of the physical properties of AlN have been reliably measured and bulk AlN synthesized.

Although GaN has been studied far more extensively than the other group III nitrides, there is still a great need for further investigations even to approach the level of understanding of technologically important materials such as Si and GaAs. GaN growth often suffers from relatively high background n-type carrier concentrations due to native defects and possibly impurities. The lack of commercially available native substrates exacerbates this situation. These, together with the difficulties in obtaining p-type doping, and the arcane fabrication processes, catalyzed the early bottlenecks, stymieing progress. Recent burgeoning interest has led to improvements in crystal-growth and processing technologies, and allowed many difficulties encountered earlier to be overcome. Consequently, a number of laboratories have begun to obtain high-quality GaN consistently with

room-temperature background electron concentrations as low as $5 \times 10^{16} \text{ cm}^{-3}$. The successful development of schemes leading to p-type GaN has led to the demonstration of excellent p-n-junction LEDs in the UV, violet, blue, green and even yellow bands of the visible spectrum with brightness suitable for outdoor displays, CW (continuous wave) lasers, power modulation doped field-effect transistors (MODFETs), and UV detectors, including those for the solar-blind region. Still, much work remains to be done in the determination of the fundamental physical properties of GaN.

InN has not received the experimental attention given to GaN and AlN. This is probably due to difficulties in growing high-quality crystalline InN samples, the poor luminescence properties of InN, and the existence of alternative, well-characterized/developed semiconductors such as AlGaAs and In (Ga, Al)AsP, which have energy band gaps close to what was thought to be the band gap of InN. More recent data and calculations appear to indicate that the band gap of InN is close to 0.7–0.8 eV. Even with this new band gap, InN does not escape competition in the sense that another well-developed semiconductor, InGaAs, covers that region. Consequently, practical applications of InN are restricted to its alloys with GaN and AlN, in addition to tandem solar cells. The growth of high-quality InN and the enumeration of its fundamental physical properties remain, for the present, a purely scientific enterprise. InN is not different from GaN and AlN in the sense that it suffers from the same lack of a suitable substrate material and, in particular, a high native-defect concentration. Moreover, early InN layers may have been polycrystalline and contained large concentrations of O and impurities and/or point defects resulting in large electron concentrations. The large band gap measured in early InN layers may have been caused by O, leading to oxides of In with large band gaps as well as a Burstein–Moss blue shift due to high electron concentrations.

References

- 32.1 H. Morkoç: *Nitride Semiconductors and Devices*, its 2nd edition will appear within 2004 (Springer, Heidelberg 1999)
- 32.2 S. Nakamura, S. Pearton, G. Fasol: *The Blue Laser Diodes* (Springer, Berlin Heidelberg New York 2001)
- 32.3 P. Ruterana, M. Albrecht, J. Neugebauer: *Nitride Semiconductors – Handbook on Materials and Devices* (Wiley, Weinheim 2003)
- 32.4 S. J. Pearton: *GaN and Related Materials* (Gordon and Breach, New York 1997)
- 32.5 B. Gil: *Group III Nitride Semiconductor Compounds* (Clarendon, Oxford 1998)
- 32.6 J. I. Pankove, T. D. Moustakas: *GaN*, Vol. 1 (Academic, New York 1998)
- 32.7 T. D. Moustakas, I. Akasaki, B. A. Monemar: III–V nitrides, Mater. Res. Soc. Symp. Proc. **449**, 482 (1997, 1998)

- 32.8 M. E. Levinshstein, S. L. Rumyantsev, M. S. Shur: *Properties of Advanced Semiconductor Materials* (Wiley, New York 2001)
- 32.9 J. H. Edgar: *Properties of Group III Nitrides* (INSPEC, London 1994)
- 32.10 J. H. Edgar, S. T. Strite, I. Akasaki, H. Amano, C. Wetzel: *Properties, Processing and Applications of Gallium Nitride and Related Semiconductors* (IN-CPEC, London 1999)
- 32.11 S. N. Mohammad, W. Kim, A. Salvador, H. Morkoc: MRS Bull. **22**, 22 (1997)
- 32.12 S. Strite, H. Morkoc: J. Vac. Sci. Technol. B **10**, 1237 (1992)
- 32.13 H. Morkoc, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, M. Burns: J. Appl. Phys. **76**, 1363 (1994)
- 32.14 S. N. Mohammad, H. Morkoc: Proc. Quant. Electr. **20**, 361 (1996)
- 32.15 H. Morkoc, S. N. Mohammad: Science **267**, 51 (1995)
- 32.16 I. Akasaki, H. Amano: Tech. Dig. Int. Electron. Devices Meet. **96**, 231 (1996)
- 32.17 S. Nakamura: MRS Bull. **22**, 29 (1997)
- 32.18 S. Nakamura: Sel. Top. Quant. Electron. **3**, 712 (1997)
- 32.19 S. Nakamura: Science **281**, 956 (1998)
- 32.20 S. J. Pearton, J. C. Zolper, R. J. Shul, F. Ren: J. Appl. Phys. **86**, 1 (1999)
- 32.21 S. J. Pearton, F. Ren, J. C. Zolper, R. J. Shul: Mater. Sci. Eng. **R30**, 55 (2000)
- 32.22 M. A. Khan, Q. Chen, J. W. Yang, C. J. Sun: Inst. Phys. Conf. Ser. **142**, 985 (1995)
- 32.23 M. A. Khan, Q. Chen, J. Yang, M. Z. Anwar, M. Blasingame, M. S. Shur: Tech. Dig. Int. Electron. Devices Meet. **96**, 27 (1996)
- 32.24 M. S. Shur, M. A. Khan: MRS Bull. **22**, 44 (1997)
- 32.25 L. Liu, J. H. Edgar: Mater. Sci. Eng. **R37**, 61 (2002)
- 32.26 S. C. Jain, M. Willander, J. Narayan, R. Van Overstraeten: J. Appl. Phys. **87**, 965 (2000)
- 32.27 B. Monemar: Semicond. Semimetals **50**, 305 (1998)
- 32.28 O. Ambacher: J. Phys. D: Appl. Phys. **31**, 2653 (1998)
- 32.29 I. Akasaki, H. Amano, I. Suemune: Inst. Phys. Conf. Ser. **142**, 7 (1996)
- 32.30 W. A. Harris: *Electronic Structure and Properties of Solids* (Dover, New York 1980) pp. 174–179
- 32.31 C. Y. Yeh, Z. W. Lu, S. Froyen, A. Zunger: Phys. Rev. B **46**, 10086 (1992)
- 32.32 G. A. Jeffery, G. S. Parry, R. L. Mozzi: J. Chem. Phys. **25**, 1024 (1956)
- 32.33 H. Schulz, K. H. Theimann: Solid State Commun. **23**, 815 (1977)
- 32.34 Q. Xia, H. Xia, A. L. Ruoff: J. Appl. Phys. **73**, 8198 (1993)
- 32.35 P. Perlin, C. Jauberthie-Carillon, J. P. Itie, A. San Miguel, I. Grzegory, A. Polian: Phys. Rev. B **45**, 83 (1992)
- 32.36 M. Ueno, M. Yoshida, A. Onodera, O. Shimomura, K. Takemura: Phys. Rev. B **49**, 14 (1994)
- 32.37 M. J. Paisley, Z. Sitar, J. B. Posthill, R. F. Davis: J. Vac. Sci. Technol. **7**, 701 (1989)
- 32.38 T. Lei, M. Fanciulli, R. J. Molnar, T. D. Moustakas, R. J. Graham, J. Scanlon: Appl. Phys. Lett. **59**, 944 (1991)
- 32.39 R. C. Powell, N. E. Lee, Y. W. Kim, J. E. Greene: J. Appl. Phys. **73**, 189 (1993)
- 32.40 M. Mizita, S. Fujieda, Y. Matsumoto, T. Kawamura: Jpn. J. Appl. Phys. **25**, L945 (1986)
- 32.41 M. Leszczynski, J. Bak-Misiuk, J. Domagala, J. Muszalski, M. Kaniewska, J. Marczewski: Appl. Phys. Lett. **67**, 539 (1995)
- 32.42 M. Leszczynski, E. Litwin-Staszewska, J. Bak-Misiuk, J. Domagala: Acta Phys. Pol. **88**, 837 (1995)
- 32.43 G. S. Cargill, A. Segmuller, T. F. Kuech, T. N. Theis: Phys. Rev. B **46**, 10078 (1992)
- 32.44 M. Leszczynski, T. Suski, P. Perlin, H. Teisseyre, I. Grzegory, M. Bockowski, J. Jun, S. Polowski, K. Pakula, J. M. Baranowski, C. T. Foxon, T. S. Cheng: Appl. Phys. Lett. **69**, 73 (1996)
- 32.45 J. Härtwing, S. Grosching: Phys. Stat. Solidi **115**, 369 (1989)
- 32.46 W. L. Bond: Acta Crystallogr. **13**, 814 (1960)
- 32.47 I. Petrov, E. Mojab, R. Powell, J. Greene, L. Hultman, J.-E. Sundgren: Appl. Phys. Lett. **60**, 2491 (1992)
- 32.48 M. E. Sherwin, T. J. Drummond: J. Appl. Phys. **69**, 8423 (1991)
- 32.49 Q. Xia, H. Xia, A. L. Ruoff: Phys. Rev. B **74**, 12925 (1993)
- 32.50 H. Vollstadt, E. Ito, M. Akaishi, S. Akimoto, O. Fukunaga: Proc. Jpn. Acad. B **66**, 7 (1990)
- 32.51 A. Munoz, K. Kunc: Phys. Rev. B **44**, 10372 (1991)
- 32.52 P. E. Van Camp, V. E. Van Doren, J. T. Devreese: Solid State Commun. **81**, 23 (1992)
- 32.53 M. Tanaka, S. Nakahata, K. Sogabe, H. Nakata, M. Tabioka: Jpn. J. Appl. Phys. **36**, L1062 (1997)
- 32.54 H. Angerer, D. Brunner, F. Freudenberger, O. Ambacher, M. Stutzmann, R. Höpler, T. Metzger, E. Born, G. Dollinger, A. Bergmaier, S. Karsch, H.-J. Körner: Appl. Phys. Lett. **71**, 1504 (1997)
- 32.55 J. Domagala, M. Leszczynski, P. Prystawko, T. Suski, R. Langer, A. Barski, M. Bremser: J. Alloy Compd. **286**, 284 (1999)
- 32.56 K. Kim, W. R. L. Lambrecht, B. Segall: Phys. Rev. B **53**, 16310 (1996). erratum: [32.388]
- 32.57 A. F. Wright, J. S. Nelson: Phys. Rev. **51**, 7866 (1995)
- 32.58 M. Leszczynski, P. Prystawko, T. Suski, B. Lucznik, J. Domagala, J. Bak-Misiuk, A. Stonert, A. Turos, R. Langer, A. Barski: J. Alloy Compd. **286**, 271 (1999)
- 32.59 T. Detchprohm, K. Hiramatsu, K. Itoh, I. Akasaki: Jpn. J. Appl. Phys. **31**, L1454 (1992)
- 32.60 M. Leszczynski, H. Teisseyre, T. Suski, I. Grzegory, M. Bockowski, J. Jun, S. Polowski, J. Major: J. Phys. D **69**, A149 (1995)
- 32.61 T. Deguchi, D. Ichiryu, K. Toshioka, K. Sekiguchi, T. Sota, R. Matsuo, T. Azuhata, M. Yamaguchi, T. Yagi, S. Chichibu, S. Nakamura: J. Appl. Phys. **86**, 1860 (1999)

- 32.62 W. Paszkowicz, J. Adamczyk, S. Krukowski, M. Leszczynski, S. Porowski, J. A. Sokolowski, M. Michalec, W. Lasocha: *Philos. Mag. A* **79**, 1145 (1999)
- 32.63 S. Strite, D. Chandrasekhar, D.J. Smith, J. Sariel, H. Chen, N. Teraguchi: *J. Cryst. Growth* **127**, 204 (1993)
- 32.64 W. Paszkowicz: *Powder Diffraction* **14**, 258 (1999)
- 32.65 F.D. Murnaghan: *Proc. Natl. Acad. Sci.* **30**, 244 (1944)
- 32.66 A. Polian, M. Grimsditch, I. Grzegory: *J. Appl. Phys.* **79**, 3343 (1996)
- 32.67 I. Yonenaga, T. Shima, M. H. F. Sluiter: *Jpn. J. Appl. Phys.* **41**, 4620 (2002)
- 32.68 I. Yonenaga: *MRS Internet J. Nitride Semicond. Res.* **7**, 6 (2002)
- 32.69 T. Tsubouchi, N. Mikoshiba: *IEEE Trans. Sonics Ultrason.* **SU 32**, 634 (1985)
- 32.70 L. E. McNeil, M. Grimsditch, R. H. French: *J. Am. Ceram. Soc.* **76**, 1132 (1993)
- 32.71 A. F. Wright: *J. Appl. Phys.* **82**, 2833 (1997)
- 32.72 T. Tsubouchi, K. Sugai, N. Mikoshiba: *Ultrasonic Symposium Proceedings* (IEEE, New York 1981) p. 375
- 32.73 R. Thokala, J. Chaudhuri: *Thin Solid Films* **266**, 189 (1995)
- 32.74 M. Ueno, A. Onodera, O. Shimomura, K. Takemura: *Phys. Rev. B* **45**, 10123 (1992)
- 32.75 D. Gerlich, S. L. Dole, G. A. Slack: *J. Phys. Chem. Solid* **47**, 437 (1986)
- 32.76 P. E. Van Camp, V. E. Van Doren, J. T. Devreese: *Phys. Rev. B* **44**, 9056 (1991)
- 32.77 E. Ruiz, S. Alvarez, P. Alemany: *Phys. Rev. B* **49**, 7617 (1994)
- 32.78 V. A. Savastenko, A. V. Sheleg: *Phys. Stat. Sol. A* **48**, K135 (1978)
- 32.79 R. B. Schwarz, K. Khachataryan, E. R. Weber: *Appl. Phys. Lett.* **70**, 1122 (1997)
- 32.80 C. Deger, E. Born, H. Angerer, O. Ambacher, M. Stutzmann, J. Hornstein, E. Riha, G. Fischer: *Appl. Phys. Lett.* **72**, 2400 (1998)
- 32.81 R. Nowak, M. Pessa, M. Suganuma, M. Leszczynski, I. Grzegory, S. Porowski, F. Yoshida: *Appl. Phys. Lett.* **75**, 2070 (1999)
- 32.82 I. Yonenaga, K. Motoki: *J. Appl. Phys.* **90**, 6539 (2001)
- 32.83 J. H. Edgar, C. H. Wei, D. T. Smith, T. J. Kistnermacher, W. A. Bryden: *J. Mater. Sci.* **8**, 307 (1997)
- 32.84 P. Perlin, A. Polian, T. Suski: *Phys. Rev. B* **47**, 2874 (1993)
- 32.85 C. T. M. Ribeiro, F. Alvarez, A. R. Zanatta: *Appl. Phys. Lett.* **81**, 1005 (2002)
- 32.86 Z. M. Ren, Y. F. Lu, H. Q. Ni, T. Y. F. Liew, B. A. Cheong, S. K. Chow, M. L. Ng, J. P. Wang: *J. Appl. Phys.* **88**, 7346 (2000)
- 32.87 O. Brafman, G. Lengyel, S. S. Mitra, P. J. Gielisse, J. N. Plendl, L. C. Mansur: *Solid State Commun.* **6**, 523 (1968)
- 32.88 K. Shimada, T. Sota, K. Suzuki: *J. Appl. Phys.* **84**, 4951 (1993)
- 32.89 K. Miwa, A. Fukumoto: *Phys. Rev. B* **48**, 7897 (1993)
- 32.90 T. Azuhata, T. Sota, K. Suzuki, S. Nakamura: *J. Phys.: Condens. Matter* **7**, L129 (1995)
- 32.91 V. Yu. Davydov, V. V. Emtsev, I. N. Goncharuk, A. N. Smirnov, V. D. Petrikov, V. V. Mamutin, V. A. Vekshin, S. V. Ivanov, M. B. Smirnov, T. Inushima: *Appl. Phys. Lett.* **75**, 3297 (1999)
- 32.92 J. S. Dyck, K. Kim, S. Limpijumnong, W. R. L. Lambrecht, K. Kash, J. C. Angus: *Solid State Commun.* **114**, 355 (2000)
- 32.93 D. Huang, F. Yun, M. A. Reshchikov, D. Wang, H. Morkoç, D. L. Rode, L. A. Farina, Ç. Kurdak, K. T. Tsen, S. S. Park, K. Y. Lee: *Solid State Electron.* **45**, 711 (2001)
- 32.94 W. Götz, N. M. Johnson, D. P. Bour, M. G. McCluskey, E. E. Haller: *Appl. Phys. Lett.* **69**, 3725 (1996)
- 32.95 C. G. Van de Walle: *Phys. Rev. B* **56**, R10020 (1997)
- 32.96 C. Wetzel, A. L. Chen, J. W. Suski, J. W. Ager III, W. Walukiewicz: *Phys. Stat. Sol. B* **198**, 243 (1996)
- 32.97 G. Kaczmarczyk, A. Kaschner, S. Reich, A. Hoffmann, C. Thomsen, D. J. As, A. P. Lima, D. Schikora, K. Lischka, R. Averbeck, H. Riechert: *Appl. Phys. Lett.* **76**, 2122 (2000)
- 32.98 K. M. Taylor, C. Lenie: *J. Electrochem. Soc.* **107**, 308 (1960)
- 32.99 V. Kirchner, H. Heinke, D. Hommel, J. Z. Domagala, M. Leszczynski: *Appl. Phys. Lett.* **77**, 1434 (2000)
- 32.100 G. A. Slack, S. F. Bartram: *J. Appl. Phys.* **46**, 89 (1975)
- 32.101 W. M. Yim, R. J. Paff: *J. Appl. Phys.* **45**, 1456 (1974)
- 32.102 H. P. Maruska, J. J. Tietjen: *Appl. Phys. Lett.* **15**, 327 (1969)
- 32.103 A. U. Sheleg, V. A. Savastenko: *Vestsi Akad. Nauk, Set. Fiz.-Mat. Nauk USSR* **3**, 126 (1976)
- 32.104 M. Leszczynski, J. F. Walker: *Appl. Phys. Lett.* **62**, 1484 (1993)
- 32.105 G. L. Slack, L. J. Schowalter, D. Morelli, J. A. Freitas Jr.: *J. Cryst. Growth* **246**, 287 (2002)
- 32.106 G. A. Slack: *J. Phys. Chem. Solids* **34**, 321 (1973)
- 32.107 G. A. Slack, T. F. McNelly: *J. Cryst. Growth* **42**, 560 (1977)
- 32.108 K. Watari, K. Ishizaki, F. Tsuchiya: *J. Mater. Sci.* **28**, 3709 (1993)
- 32.109 A. Witek: *Diamond Relat. Mater.* **7**, 962 (1998)
- 32.110 A. Nikolaev, I. Nikitina, A. Zubrilov, M. Mynbaeva, Y. Melnik, V. Dmitriev: *Mater. Res. Soc. Symp. Proc.* **595**, 6.5.1 (2000)
- 32.111 D. I. Florescu, V. M. Asnin, F. H. Pollak: *Compound Semiconductor* **7**, 62 (2001)
- 32.112 E. K. Sichel, J. I. Pankove: *J. Phys. Chem. Solids* **38**, 330 (1977)
- 32.113 D. Kotchetkov, J. Zou, A. A. Balandin, D. I. Florescu, F. H. Pollak: *Appl. Phys. Lett.* **79**, 4316 (2001)
- 32.114 J. Zou, D. Kotchetkov, A. A. Balandin, D. I. Florescu, F. H. Pollak: *J. Appl. Phys.* **92**, 2534 (2002)
- 32.115 D. I. Florescu, V. M. Asnin, F. H. Pollak, R. J. Molnar: *Mater. Res. Soc. Symp. Proc.* **595**, 3.89.1 (2000)

- 32.116 D. I. Florescu, V. M. Asnin, F. H. Pollak, R. J. Molnar, C. E. C. Wood: *J. Appl. Phys.* **88**, 3295 (2000)
- 32.117 G. A. Slack, R. A. Tanzilli, R. O. Pohl, J. W. Vandersande: *J. Phys. Chem. Solids* **48**, 641 (1987)
- 32.118 S. Krukowski, A. Witek, J. Adamczyk, J. Jun, M. Bockowski, I. Grzegory, B. Lucznik, G. Nowak, M. Wroblewski, A. Presz, S. Gierlotka, S. Stelmach, B. Palosz, S. Porowski, P. Zinn: *J. Phys. Chem. Solids* **59**, 289 (1998)
- 32.119 A. D. Mah, E. G. King, W. W. Weller, A. U. Christensen: *Bur. Mines, Rept. Invest.* **RI-5716**, 18 (1961)
- 32.120 V. P. Glushko, L. V. Gurevich, G. A. Bergman, I. V. Weitz, V. A. Medvedev, G. A. Chachkurov, V. S. and Yungman: *Termodinamicheskie Svoistwa Individualnykh Weshchestw (the old USSR)*, Vol. 1 (Nauka, Moscow 1979) p. 164
- 32.121 I. Basin, O. Knacke, O. Kubaschewski: *Thermochemical Properties of Inorganic Substances* (Springer, Berlin, Heidelberg 1977)
- 32.122 V. I. Koshchenko, A. F. Demidienko, L. D. Sabanova, V. E. Yachmenev, V. E. Gran, A. E. Radchenko: *Inorg. Mater.* **15**, 1329 (1979)
- 32.123 S. Krukowski, M. Leszczynski, S. Porowski: Thermal properties of the Group III nitrides. In: *Properties, Processing and Applications of Gallium Nitride and Related Semiconductors*, EMIS Datareviews Series, No. 23, ed. by J. H. Edgar, S. Strite, I. Akasaki, H. Amano, C. Wetzel (INSPEC, The Institution of Electrical Engineers, Stevenage, UK 1999) p. 23
- 32.124 A. F. Demidienko, V. I. Koshchenko, L. D. Sabanova, V. E. Gran: *Russ. J. Phys. Chem.* **49**, 1585 (1975)
- 32.125 J. C. Nipko, C.-K. Loong, C. M. Balkas, R. F. Davis: *Appl. Phys. Lett.* **73**, 34 (1998)
- 32.126 V. Davydov, A. Klochikhin, S. Ivanov, J. Aderhold, A. Yamamoto: Growth and properties of InN. In: *Nitride Semiconductors – Handbook on Materials and Devices*, ed. by P. Ruterana, M. Albrecht, J. Neugebauer (Wiley, New York 2003)
- 32.127 D. L. Rode, R. K. Willardson, A. C. Beer (Eds.): *Semiconductors and Semimetals*, Vol. 10 (Academic, New York 1975) pp. 1–90
- 32.128 K. Seeger: *Semiconductor Physics*, 2 edn. (Springer, Berlin Heidelberg New York 1982)
- 32.129 V. W. L. Chin, T. L. Tansley, T. Osotchan: *J. Appl. Phys.* **75**, 7365 (1994)
- 32.130 D. L. Rode, D. K. Gaskill: *Appl. Phys. Lett.* **66**, 1972 (1995)
- 32.131 M. E. Lin, B. Sverdlov, G. L. Zhou, H. Morkoç: *Appl. Phys. Lett.* **62**, 3479 (1993)
- 32.132 H. M. Ng, D. Doppalapudi, T. D. Moustakas, N. G. Weimann, L. F. Eastman: *Appl. Phys. Lett.* **73**, 821 (1998)
- 32.133 N. G. Weimann, L. F. Eastman, D. Doppalapudi, H. M. Ng, T. D. Moustakas: *J. Appl. Phys.* **83**, 3656 (1998)
- 32.134 D. C. Look, J. R. Sizelove: *Phys. Rev. Lett.* **82**, 1237 (1999)
- 32.135 H. W. Choi, J. Zhang, S. J. Chua: *Mater. Sci. Semi-cond. Process.* **4**, 567 (2001)
- 32.136 J. Y. Shi, L. P. Yu, Y. Z. Wang, G. Y. Zhang, H. Zhang: *Appl. Phys. Lett.* **80**, 2293 (2002)
- 32.137 Z. Q. Fang, D. C. Look, W. Kim, Z. Fan, A. Botchkarev, H. Morkoc: *Appl. Phys. Lett.* **72**, 2277 (1998)
- 32.138 K. Wook, A. E. Botchkarev, H. Morkoc, Z. Q. Fang, D. C. Look, D. J. Smith: *J. Appl. Phys.* **84**, 6680 (1998)
- 32.139 Q. S. Zhu, N. Sawaki: *Appl. Phys. Lett.* **76**, 1594 (2000)
- 32.140 Z. Chen, Y. Yuan, Da-C. Lu, X. Sun, S. Wan, X. Liu, P. Han, X. Wang, Q. Zhu, Z. Wang: *Solid State Electron.* **46**, 2069 (2002)
- 32.141 S. Nakamura: *Jpn. J. Appl. Phys.* **30**, L1705 (1991)
- 32.142 W. Götz, L. T. Romano, J. Walker, N. M. Johnson, R. J. Molnar: *Appl. Phys. Lett.* **72**, 1214 (1998)
- 32.143 D. C. Look, R. J. Molnar: *Appl. Phys. Lett.* **70**, 3377 (1997)
- 32.144 S. Nakamura, T. Mukai, M. Senoh: *Jpn. J. Appl. Phys.* **31**, 2883 (1992)
- 32.145 S. Nakamura, T. Mukai, M. Senoh: *J. Appl. Phys.* **71**, 5543 (1992)
- 32.146 S. C. Binari, H. C. Dietrich: In: *GaN and Related Materials*, ed. by S. J. Pearton (Gordon and Breach, New York 1997) pp. 509–534
- 32.147 F. Yun, M. A. Reshchikov, K. Jones, P. Visconti, S. S. Park, K. Y. Lee: *Solid State Electron.* **44**, 2225 (2000)
- 32.148 B. Heying, I. Smorchkova, C. Poblenz, C. Elsass, P. Fini, S. Den Baars, U. Mishra, J. S. Speck: *Appl. Phys. Lett.* **77**, 2885 (2000)
- 32.149 D. D. Koleske, R. L. Henry, M. E. Twigg, J. C. Culbertson, S. C. Binari, A. E. Wickenden, M. Fatemi: *Appl. Phys. Lett.* **80**, 4372 (2000)
- 32.150 S. Nakamura, T. Mukai: *Jpn. J. Appl. Phys.* **31**, L1457 (1992)
- 32.151 W. Götz, N. M. Johnson, C. Chen, H. Liu, C. Kuo, W. Imler: *Appl. Phys. Lett.* **68**, 3144 (1996)
- 32.152 M. Ilegams, H. C. Montgomery: *J. Phys. Chem. Solids* **34**, 885 (1973)
- 32.153 M. A. Littlejohn, J. R. Hauser, M. Glisson: *Appl. Phys. Lett.* **26**, 625 (1975)
- 32.154 U. V. Bhapkar, M. S. Shur: *J. Appl. Phys.* **82**, 1649 (1997)
- 32.155 J. D. Albrecht, R. P. Wang, P. P. Ruden, M. Farahmand, K. F. Brennan: *J. Appl. Phys.* **83**, 1446 (1998)
- 32.156 S. Dhar, S. Ghosh: *J. Appl. Phys.* **86**, 2668 (1999)
- 32.157 J. G. Kim, A. C. Frenkel, H. Liu, R. M. Park: *Appl. Phys. Lett.* **65**, 91 (1994)
- 32.158 R. J. Molnar, W. Götz, L. T. Romano, N. M. Johnson: *J. Cryst. Growth* **178**, 147 (1997)
- 32.159 S. Yamaguchi, M. Kariya, S. Nitta, T. Takeuchi, C. Wetzel, H. Amano, I. Akasaki: *J. Appl. Phys.* **85**, 7682 (1999)
- 32.160 H. Morkoç: *IEEE J. Select. Top. Quant. Electron.* **4**, 537 (1998)

- 32.161 D. C. Look, D. C. Reynolds, J. W. Hemsky, J. R. Sizemore, R. L. Jones, R. J. Molnar: *Phys. Rev. Lett.* **79**, 2273 (1997)
- 32.162 S. Keller, B. P. Keller, Y. F. Wu, B. Heying, D. Kapolnek, J. S. Speck, U. K. Mishra, S. P. Den Baars: *Appl. Phys. Lett.* **68**, 1525 (1996)
- 32.163 S. Nakamura, T. Mukhai, M. Senoh: *J. Appl. Phys.* **75**, 7365 (1997)
- 32.164 M. A. Khan, R. A. Skogman, R. G. Schulze, M. Ger-shenzon: *Appl. Phys. Lett.* **42**, 430 (1983)
- 32.165 I. Akasaki, H. Amano: *Mater. Res. Soc. Symp. Proc.* **242**, 383 (1992)
- 32.166 T. Matsuoka: *Mater. Res. Soc. Symp. Proc.* **395**, 39 (1995)
- 32.167 S. Sinharoy, A. K. Aggarwal, G. Augustine, L. B. Rawland, R. L. Messham, M. C. Driver, R. H. Hopkins: *Mater. Res. Soc. Symp. Proc.* **395**, 157 (1995)
- 32.168 R. J. Molnar, R. Aggarwal, Z. L. Lian, E. R. Brown, I. Melngailis, W. Götz, L. T. Romano, N. M. Johnson: *Mater. Res. Soc. Symp. Proc.* **395**, 157 (1995)
- 32.169 T. T. Mnatsakanov, M. E. Levinshstein, L. I. Pomortseva, S. N. Yurkov, G. S. Simin, M. A. Khan: *Solid State Electron.* **47**, 111 (2003)
- 32.170 D. M. Caughey, R. E. Thomas: *Proc. IEEE* **55**, 2192 (1967)
- 32.171 J. Edwards, K. Kawabe, G. Stevens, R. H. Tredgold: *Solid State Commun.* **3**, 99 (1965)
- 32.172 K. Kawabe, R. H. Tredgold, Y. Inyishi: *Electr. Eng. Jpn.* **87**, 62 (1967)
- 32.173 G. A. Cox, D. O. Cummins, K. Kawabe, R. H. Tredgold: *J. Phys. Chem. Solids* **28**, 543 (1967)
- 32.174 W. M. Yim, E. J. Stotko, P. J. Zanzucchi, J. Pankove, M. Ettenberg, S. L. Gilbert: *J. Appl. Phys.* **44**, 292 (1973)
- 32.175 S. Yoshida, S. Misawa, Y. Fujii, S. Takada, H. Hayakawa, S. Gonda, A. Itoh: *J. Vac. Sci. Technol.* **16**, 990 (1979)
- 32.176 A. G. Gorbato, V. M. Kamysnoc: *Sov. Powder, Metall. Met. Ceram.* **9**, 917 (1970)
- 32.177 R. F. Rutz: *Appl. Phys. Lett.* **28**, 379 (1976)
- 32.178 R. F. Rutz, E. P. Harrison, J. J. Cuome: *IBM J. Res. Dev.* **17**, 61 (1973)
- 32.179 A. Fara, F. Bernadini, V. Fiorentini: *J. Appl. Phys.* **85**, 2001 (1999)
- 32.180 K. Wongchotigul, N. Chen, D. P. Zhang, X. Tang, M. G. Spencer: *Mater. Lett.* **26**, 223 (1996)
- 32.181 H. J. Hovel, J. J. Cuomo: *Appl. Phys. Lett.* **20**, 71 (1972)
- 32.182 T. L. Tansley, C. P. Foley: *Electron. Lett.* **20**, 1066 (1984)
- 32.183 M. J. Brett, K. L. Westra: *Thin Solid Films* **192**, 227 (1990)
- 32.184 J. S. Ergan, T. J. Kistenmacher, W. A. Bryden, S. A. Ecelberger: *Proc. Mater. Res. Soc.* **202**, 383 (1991)
- 32.185 T. J. Kistenmacher, W. A. Bryden: *Appl. Phys. Lett.* **59**, 1844 (1991)
- 32.186 C. R. Abernathy, S. J. Pearton, F. Ren, P. W. Wisk: *J. Vac. Sci. Technol. B* **11**, 179 (1993)
- 32.187 W. R. Bryden, S. A. Ecelberger, M. E. Hawley, T. J. Kistenmacher: *MRS Proc.* **339**, 497 (1994)
- 32.188 T. Maruyama, T. Morishita: *J. Appl. Phys.* **76**, 5809 (1994)
- 32.189 H. Lu, W. J. Schaff, J. Hwang, H. Wu, W. Yeo, A. Pharkya, L. F. Eastman: *Appl. Phys. Lett.* **77**, 2548 (2000)
- 32.190 Y. Saito, N. Teraguchi, A. Suzuki, T. Araki, Y. Nanishi: *Jpn. J. Appl. Phys.* **40**, L91 (2001)
- 32.191 J. Aderhold, V. Yu. Davydov, F. Fedler, H. Klausling, D. Miste, T. Rotter, O. Semchinova, J. Stemmer, J. Graul: *J. Cryst. Growth* **221**, 701 (2001)
- 32.192 H. Lu, W. J. Schaff, J. Hwang, H. Wu, G. Koley, L. F. Eastman: *Appl. Phys. Lett.* **79**, 1489 (2001)
- 32.193 Motlan, E. M. Goldys, T. L. Tansley: *J. Cryst. Growth* **241**, 165 (2002)
- 32.194 Y. Saito, T. Yamaguchi, H. Kanazawa, K. Kano, T. Araki, Y. Nanishi, N. Teraguchi, A. Suzuki: *J. Cryst. Growth* **237–239**, 1017 (2002)
- 32.195 M. Higashiwaki, T. Matsui: *Jpn. J. Appl. Phys.* **41**, L540 (2002)
- 32.196 H. W. Lu, J. Schaff, L. F. Eastman, J. Wu, W. Walukiewicz, K. M. Yu, J. W. Auger III, E. E. Haller, O. Ambacher: *Conference Digest of the 44th Electronic Materials Conference*, Santa Barbara, p. 2 (2002)
- 32.197 M. Higashiwaki, T. Matsui: *J. Cryst. Growth* **252**, 128 (2003)
- 32.198 T. L. Tansley, C. P. Foley, J. S. Blakemore (Ed.): *Proc. 3rd Int. Conf. on Semiinsulating III–V Materials*, Warm Springs, OR 1984 (Shiva, London 1985)
- 32.199 T. L. Tansley, R. J. Egan, E. C. Horrigan: *Thin Solid Films* **164**, 441 (1988)
- 32.200 C. Stampfl, C. G. Van de Walle, D. Vogel, P. Kruger, J. Pollmann: *Phys. Rev. B* **61**, 7846 (2000)
- 32.201 D. C. Look, H. Lu, W. J. Schaff, J. Jasinski, Z. Liliental-Weber: *Appl. Phys. Lett.* **80**, 258 (2002)
- 32.202 B. E. Foutz, S. K. O'Leary, M. S. Shur, L. F. Eastman: *J. Appl. Phys.* **85**, 7727 (1999)
- 32.203 J. D. Albrecht, R. P. Wang, P. P. Ruden, M. Farahmand, K. F. Brennan: *J. Appl. Phys.* **83**, 4777 (1998)
- 32.204 S. K. O'Leary, B. E. Foutz, M. S. Shur, U. V. Bhapkar, L. F. Eastman: *Solid State Commun.* **105**, 621 (1998)
- 32.205 B. Gelmont, K. Kim, M. Shur: *J. Appl. Phys.* **74**, 1818 (1993)
- 32.206 N. S. Mansour, K. W. Kim, M. A. Littlejohn: *J. Appl. Phys.* **77**, 2834 (1995)
- 32.207 J. Kolnik, I. H. Oguzman, K. F. Brennan, R. Wang, P. P. Ruden, Y. Wang: *J. Appl. Phys.* **78**, 1033 (1995)
- 32.208 M. Shur, B. Gelmont, M. A. Khan: *J. Electron. Mater.* **25**, 777 (1996)
- 32.209 S. K. O'Leary, B. E. Foutz, M. S. Shur, U. V. Bhapkar, L. F. Eastman: *J. Appl. Phys.* **83**, 826 (1998)
- 32.210 E. Bellotti, B. K. Doshi, K. F. Brennan, J. D. Albrecht, P. P. Ruden: *J. Appl. Phys.* **85**, 916 (1999)

- 32.211 C. G. Rodrigues, V. N. Freire, A. R. Vasconcellos, R. Luzzi: Appl. Phys. Lett. **76**, 1893 (2000)
- 32.212 M. Wraback, H. Shen, S. Rudin, E. Bellotti: Phys. Stat. Sol. (b) **234**, 810 (2002)
- 32.213 R. Collazo, R. Schesser, Z. Sitar: Appl. Phys. Lett. **81**, 5189 (2002)
- 32.214 B. Monemar, U. Lindefelt, W. M. Chen: Physica B **146**, 256 (1987)
- 32.215 M. Tchoukueu, O. Briot, B. Gil, J. P. Alexis, R. L. Aulombard: J. Appl. Phys. **80**, 5352 (1996)
- 32.216 R. Dingle, D. D. Sell, S. E. Stokowski, M. Ilegems: Phys. Rev. B **4**, 1211 (1971)
- 32.217 E. L. Ivchenko: *Excitons* (North Holland, Amsterdam 1982) p. 141
- 32.218 B. Monemar, J. P. Bergman, I. A. Buyanova: Optical characterization of GaN and related materials. In: *GaN and Related Material*, ed. by S. J. Pearton (Golden and Breach, Amsterdam 1997) p. 85
- 32.219 J. J. Song, W. Shan: In: *Group III Nitride Semiconductor Compounds*, ed. by B. Gil (Clarendon, Oxford 1998) pp. 182–241
- 32.220 B. Monemar, J. P. Bergman, I. A. Buyanova: In: *GaN and Related Materials Semiconductor Compounds*, ed. by S. J. Pearton (Gordon and Breach, New York 1998) pp. 85–139
- 32.221 B. Monemar: Phys. Rev. B **10**, 676 (1974)
- 32.222 B. Gil, O. Briot, R. L. Aulombard: Phys. Rev. B **52**, R17028 (1995)
- 32.223 W. Shan, B. D. Little, A. J. Fischer, J. J. Song, B. Goldenberg, W. G. Perry, M. D. Bremser, R. F. Davis: Phys. Rev. B **54**, 16369 (1996)
- 32.224 M. Smith, G. D. Chen, J. Y. Lin, H. X. Jiang, A. Salvador, W. K. Kim, O. Aktas, A. Botchkarev, H. Morkoç: Appl. Phys. Lett. **67**, 3387 (1995)
- 32.225 K. P. Korona, A. Wyszomolek, K. Pakula, R. Stepniewski, J. M. Baranowski, I. Grzegory, B. Lucznik, M. Wroblewski, S. Porowski: Appl. Phys. Lett. **69**, 788 (1996)
- 32.226 K. Kornitzer, T. Ebner, M. Grehl, K. Thonke, R. Sauer, C. Kirchner, V. Schwegler, M. Kamp, M. Leszczynski, I. Grzegory, S. Porowski: Phys. Stat. Sol. (b) **216**, 5 (1999)
- 32.227 B. J. Skromme, K. Palle, C. D. Poweleit, H. Yamane, M. Aoki, F. J. Disalvo: J. Cryst. Growth **246**, 299 (2002)
- 32.228 L. Eckey, L. Podlosowski, A. Goldner, A. Hoffmann, I. Broser, B. K. Meyer, D. Volm, T. Streibl, K. Hiramatsu, T. Detcpohm, H. Amano, I. Akasaki: Ins. Phys. Conf. Ser. **142**, 943 (1996)
- 32.229 K. Pakula, A. Wyszomolek, K. P. Korona, J. M. Baranowski, R. Stepniewski, I. Grzegory, M. Bockowski, J. Jun, S. Krukowski, M. Wroblewski, S. Porowski: Solid State Commun. **97**, 919 (1996)
- 32.230 A. V. Rodina, M. Dietrich, A. Goldner, L. Eckey, A. L. L. Efros, M. Rosen, A. Hoffmann, B. K. Meyer: Phys. Stat. Sol. (b) **216**, 216 (1999)
- 32.231 A. Shikanai, T. Azuhata, T. Sota, S. Chichibu, A. Kuramata, K. Horino, S. Nakamura: J. Appl. Phys. **81**, 417 (1997)
- 32.232 B. Monemar, J. P. Bergman, I. A. Buyanova, W. Li, H. Amano, I. Akasaki: MRS Int. J. Nitride Semicond. Res. **1**, 2 (1996)
- 32.233 S. F. Chichibu, K. Torii, T. Deguchi, T. Sota, A. Setoguchi, H. Nakanishi, T. Azuhata, S. Nakamura: Appl. Phys. Lett. **76**, 1576 (2000)
- 32.234 J. F. Muth, J. H. Lee, I. K. Shmagin, R. M. Kolbas, H. C. Casey Jr., B. P. Keller, U. K. Mishra, S. P. DenBaars: Appl. Phys. Lett. **71**, 2572 (1997)
- 32.235 R. Stepniewski, K. P. Korona, A. Wyszomolek, J. M. Baranowski, K. Pakula, M. Potemski, G. Martinez, I. Grzegory, S. Porowski: Phys. Rev. B **56**, 15151 (1997)
- 32.236 W. Shan, A. J. Fischer, S. J. Hwang, B. D. Little, R. J. Hauenstein, X. C. Xie, J. J. Song, D. S. Kim, B. Goldenberg, R. Horning, S. Krishnankutty, W. G. Perry, M. D. Bremser, R. F. Davis: J. Appl. Phys. **83**, 455 (1998)
- 32.237 B. Gil, S. Clur, O. Briot: Solid State Commun. **104**, 267 (1997)
- 32.238 W. Shan, R. J. Hauenstein, A. J. Fischer, J. J. Song, W. G. Perry, M. D. Bremser, R. F. Davis, B. Goldenberg: Appl. Phys. Lett. **66**, 985 (1995)
- 32.239 C. F. Li, Y. S. Huang, L. Malikova, F. H. Pollak: Phys. Rev. B **55**, 9251 (1997)
- 32.240 A. K. Viswanath, J. I. Lee, D. Kim, C. R. Lee, J. Y. Leam: Phys. Rev. B **58**, 16333 (1998)
- 32.241 S. Chichibu, T. Azuhata, T. Sota, S. Nakamura: J. Appl. Phys. **79**, 2784 (1996)
- 32.242 M. Smith, G. D. Chen, J. Y. Lin, H. X. Jiang, M. A. Khan, C. J. Sun, Q. Chen, J. W. Yang: J. Appl. Phys. **79**, 7001 (1996)
- 32.243 M. Smith, G. D. Chen, J. Y. Lin, H. X. Jiang, A. Salvador, B. N. Sverdlov, A. Botchkarev, H. Morkoç: Appl. Phys. Lett. **66**, 3474 (1995)
- 32.244 F. Hamdani, A. Botchkarev, H. Tang, W. K. Kim, H. Morkoç: Appl. Phys. Lett. **71**, 3111 (1997)
- 32.245 F. Hamdani, A. Botchkarev, W. Kim, H. Morkoç, M. Yeadon, J. M. Gibson, S. C. Y. Tsen, D. J. Smith, D. C. Reynolds, D. C. Look, K. Evans, C. W. Litton, W. C. Mitchel, P. Hemenger: Appl. Phys. Lett. **70**, 467 (1997)
- 32.246 W. Shan, A. J. Fischer, J. J. Song, G. E. Bulman, H. S. Kong, M. T. Leonard, W. G. Perry, M. D. Bremser, B. Goldenberg, R. F. Davis: Appl. Phys. Lett. **69**, 740 (1996)
- 32.247 S. Chichibu, T. Azuhata, T. Sota, H. Amano, I. Akasaki: Appl. Phys. Lett. **70**, 2085 (1997)
- 32.248 I. A. Buyanova, J. P. Bergman, B. Monemar, H. Amano, I. Akasaki: Appl. Phys. Lett. **69**, 1255 (1996)
- 32.249 B. J. Skromme, K. C. Palle, C. D. Poweleit, H. Yamane, M. Aoki, F. J. Disalvo: Appl. Phys. Lett. **81**, 3765 (2002)

- 32.250 Y. S. Huang, Fred H. Pollak, S. S. Park, K. Y. Lee, H. Morkoç: *J. Appl. Phys.* **94**, 899 (2003)
- 32.251 B. K. Meyer: In: *Free and Bound Excitons in GaN Epitaxial Films*, MRS Proc., Vol. 449, ed. by F. A. Ponce, T. D. Moustakas, I. Akasaki, B. A. Monemar (Materials Research Society, Pittsburgh, Pennsylvania 1997) p. 497
- 32.252 B. Monemar, J. P. Bergman, I. A. Buyanova, H. Amano, I. Akasaki, T. Detchprohm, K. Hiramatsu, N. Sawaki: *Solid State Electron.* **41**, 239 (1995)
- 32.253 D. Kovalev, B. Averboukh, D. Volm, B. K. Meyer: *Phys. Rev. B* **54**, 2518 (1996)
- 32.254 G. D. Chen, M. Smith, J. Y. Lin, H. X. Jiang, S.-H. Wei, M. A. Khan, C. J. Sun: *Appl. Phys. Lett.* **68**, 2784 (1996)
- 32.255 G. Mahler, U. Schroder: *Phys. Stat. Sol. (b)* **61**, 629 (1974)
- 32.256 B. Monemar: In: *Gallium Nitride I*, ed. by J. I. Pankove, T. D. Moustakas (Academic, San Diego 1998) p. 305
- 32.257 D. C. Reynolds, D. C. Look, W. Kim, O. Aktas, A. Botchkarev, A. Salvador, H. Morkoç, D. N. Talwar: *J. Appl. Phys.* **80**, 594 (1996)
- 32.258 R. Stepniewski, M. Potemski, A. Wyszomolek, K. Pakula, J. M. Baranowski, J. Lusakowski, I. Grzegory, S. Porowski, G. Martinez, P. Wyder: *Phys. Rev. B* **60**, 4438 (1999)
- 32.259 G. Neu, M. Teisseire, E. Frayssinet, W. Knap, M. L. Sadowski, A. M. Witowski, K. Pakula, M. Leszczynski, P. Prystawko: *Appl. Phys. Lett.* **77**, 1348 (2000)
- 32.260 J. M. Baranowski, Z. Liliental-Weber, K. Korona, K. Pakula, R. Stepniewski, A. Wyszomolek, I. Grzegory, G. Nowak, S. Porowski, B. Monemar, P. Bergman: *III-V Nitrides*, Vol. 449 (MRS Proc., Pittsburg, PA 1997) p. 393
- 32.261 J. M. Baranowski and S. Porowski, *Proc. 23rd Int. Conf. on Physics of Semiconductors*, Berlin, p. 497 (1996)
- 32.262 S. J. Pearton, C. R. Abernathy, J. W. Lee, C. B. Vartuli, C. B. Mackenzi, J. D. Ren, R. G. Wilson, J. M. Zavada, R. J. Shul, J. C. Zolper: *Mater. Res. Soc. Symp. Proc.* **423**, 124 (1996)
- 32.263 B. Monemar: *J. Mater. Sci.: Mater. Electron.* **10**, 227 (1999)
- 32.264 M. Reshchikov, D. Huang, F. Yun, L. He, H. Morkoç, D. C. Reynolds, S. S. Park, K. Y. Lee: *Appl. Phys. Lett.* **79**, 3779 (2001)
- 32.265 A. K. Viswanath: *Semicond. Semimetals* **73**, 63 (2002)
- 32.266 B. Monemar, P. P. Paskov, T. Paskova, J. P. Bergman, G. Pozina, W. M. Chan, P. N. Hai, I. A. Buyanova, H. Amano, I. Akasaki: *Mater. Sci. Eng. B* **93**, 112 (2002)
- 32.267 G. Pozina, N. V. Edwards, J. P. Bergman, T. Paskova, B. Monemar, M. D. Bremser, R. F. Davis: *Appl. Phys. Lett.* **78**, 1062 (2001)
- 32.268 S. Pau, J. Kuhl, M. A. Khan, C. J. Sun: *Phys. Rev. B* **58**, 12916 (1998)
- 32.269 G. Pozina, J. P. Bergman, T. Paskova, B. Monemar: *Appl. Phys. Lett.* **75**, 412 (1999)
- 32.270 M. Leroux, B. Beaumont, N. Grandjean, P. Lorenzini, S. Haffouz, P. Vennegues, J. Massies, P. Gibart: *Mater. Sci. Eng. B* **50**, 97 (1997)
- 32.271 M. Leroux, N. Grandjean, B. Beaumont, G. Nataf, F. Semond, J. Massies, P. Gibart: *J. Appl. Phys.* **86**, 3721 (1999)
- 32.272 B. J. Skromme, G. L. Martinez: *Mater. Res. Soc. Symp.* **595**, W9.8. (1999)
- 32.273 D. C. Reynolds, D. C. Look, B. Jogai, V. M. Phanse, R. P. Vaudo: *Solid State Commun.* **103**, 533 (1997)
- 32.274 B. Santic, C. Merz, U. Kaufmann, R. Niebuhr, H. Obloh, K. Bachem: *Appl. Phys. Lett.* **71**, 1837 (1997)
- 32.275 A. K. Viswanath, J. I. Lee, S. Yu, D. Kim, Y. Choi, C. H. Hong: *J. Appl. Phys.* **84**, 3848 (1998)
- 32.276 R. A. Mair, J. Li, S. K. Duan, J. Y. Lin, H. X. Jiang: *Appl. Phys. Lett.* **74**, 513 (1999)
- 32.277 D. G. Thomas, J. J. Hopfield: *Phys. Rev.* **128**, 2135 (1962)
- 32.278 H. Saito, S. Shionoya, E. Hanamura: *Solid State Commun.* **12**, 227 (1973)
- 32.279 U. Kaufmann, M. Kunzer, C. Merz, I. Akasaki, H. Amano: *Mater. Res. Soc. Symp. Proc.* **395**, 633 (1996)
- 32.280 K. P. Korona, J. P. Bergman, B. Monemar, J. M. Baranowski, K. Pakula, L. Grzegory, S. Porowski: *Mater. Sci. Forum* **258-263**, 1125 (1997)
- 32.281 G. Neu, M. Teisseire, N. Grandjean, H. Lahreche, B. Beaumont, I. Grzegory, S. and Porowski: *Proc. Phys.* **87**, 1577 (2001)
- 32.282 C. H. Henry, K. Nassau: *Phys. Rev. B* **1**, 1628 (1970)
- 32.283 O. Lagerstedt, B. Monemar: *J. Appl. Phys.* **45**, 2266 (1974)
- 32.284 R. Dingle, M. Ilegems: *Solid State Commun.* **9**, 175 (1971)
- 32.285 S. Fischer, C. Wetzel, E. E. Haller, B. K. Meyer: *Appl. Phys. Lett.* **67**, 1298 (1995)
- 32.286 M. A. L. Johnson, Z. Yu, C. Boney, W. C. Hughes, J. W. Cook Jr, J. F. Schetzina, H. Zao, B. J. Skromme, J. A. Edmond: *MRS Proc.* **449**, 271 (1997)
- 32.287 M. Smith, G. D. Chen, J. Y. Lin, H. X. Jiang, A. Salvador, B. N. Sverdlov, A. Botchkarev, H. Morkoc, B. Goldenberg: *Appl. Phys. Lett.* **68**, 1883 (1996)
- 32.288 G. D. Chen, M. Smith, J. Y. Lin, H. X. Jiang, A. Salvador, B. N. Sverdlov, A. Botchkarev, H. Morkoc: *J. Appl. Phys.* **79**, 2675 (1995)
- 32.289 T. Ogino, M. Aoki: *Jpn. J. Appl. Phys.* **19**, 2395 (1980)
- 32.290 K. Saarinen, T. Laine, S. Kuisma, J. Nissilä, P. Hauttojärvi, L. Dobrzynski, J. M. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wyszomolek, T. Suski, M. Leszczynski, I. Grzegory, S. Porowski: *Phys. Rev. Lett.* **79**, 3030 (1997)

- 32.291 J. Neugebauer, C. G. Van de Walle: Appl. Phys. Lett. **69**, 503 (1996)
- 32.292 T. Mattila, R. M. Nieminen: Phys. Rev. **55**, 9571 (1997)
- 32.293 E. R. Glaser, T. A. Kennedy, K. Doverspike, L. B. Rowland, D. K. Gaskill, J. A. Freitas Jr, M. Asif Khan, D. T. Olson, J. N. Kuznia, D. K. Wickenden: Phys. Rev. B **51**, 13326 (1995)
- 32.294 P. Perlin, T. Suski, H. Teisseyre, M. Leszczynski, I. Grzegory, J. Jun, S. Porowski, P. Boguslawski, J. Berholc, J. C. Chervin, A. Polian, T. D. Moustakas: Phys. Rev. Lett. **75**, 296 (1995)
- 32.295 R. Zhang, T. F. Kuech: Appl. Phys. Lett. **72**, 1611 (1998)
- 32.296 I. Shalish, L. Kronik, G. Segal, Y. Rosenwaks, Y. Shapira, U. Tisch, J. Salzman: Phys. Rev. B **59**, 9748 (1999)
- 32.297 E. Calleja, F. J. Sanchez, D. Basak, M. A. Sanchez-Garsia, E. Munoz, I. Izpura, F. Calle, J. M. G. Tijero, J. L. Sanchez-Rojas, B. Beaumont, P. Lorenzini, P. Gibart: Phys. Rev. B **55**, 4689 (1997)
- 32.298 M. A. Reshchikov, F. Shahedipour, R. Y. Korotkov, M. P. Ulmer, B. W. Wessels: Physica B **273–274**, 103 (1999)
- 32.299 J. Elsner, R. Jones, M. I. Heggie, P. K. Sitch, M. Haugk, Th. Frauenheim, S. Öberg, P. R. Briddon: Phys. Rev. B **58**, 12571 (1998)
- 32.300 F. A. Ponce, D. P. Bour, W. Gotz, P. J. Wright: Appl. Phys. Lett. **68**, 57 (1996)
- 32.301 D. M. Hofmann, D. Kovalev, G. Steude, B. K. Meyer, A. Hoffmann, L. Eckey, R. Heitz, T. Detchprom, H. Amano, I. Akasaki: Phys. Rev. B **52**, 16702 (1995)
- 32.302 R. Y. Korotkov, M. A. Reshchikov, B. W. Wessels: Physica B **273–274**, 80 (1999)
- 32.303 D. G. Thomas, J. J. Hopfield, W. M. Augustyniak: Phys. Rev. A **140**, 202 (1965)
- 32.304 R. Seitz, C. Gaspar, T. Monteiro, E. Pereira, M. Leroux, B. Beaumont, P. Gibart: MRS Internet J. Nitride Semicond. Res. **2**, article 36 (1997)
- 32.305 M. Godlewski, V. Yu. Ivanov, A. Kaminska, H. Y. Zuo, E. M. Goldys, T. L. Tansley, A. Barski, U. Rossner, J. L. Rouvire, M. Arlery, I. Grzegory, T. Suski, S. Porowski, J. P. Bergman, B. Monemar: Mat. Sci. Forum **258–263**, 1149 (1997)
- 32.306 H. Haag, B. Hönerlage, O. Briot, R. L. Aulombard: Phys. Rev. B **60**, 11624 (1999)
- 32.307 M. A. Reshchikov, F. Yun, H. Morkoç, S. S. Park, K. Y. Lee: Appl. Phys. Lett. **78**, 2882 (2001)
- 32.308 S. Pacesova, L. Jastrabik: Czech. J. Phys. **B 29**, 913 (1979)
- 32.309 R. A. Youngman, J. H. Harris: J. Am. Ceram. Soc. **73**, 3238 (1990)
- 32.310 J. H. Harris, R. A. Youngman, R. G. Teller: J. Mater. Res. **5**, 1763 (1990)
- 32.311 M. Katsikini, E. C. Paloura, T. S. Cheng, C. T. Foxon: J. Appl. Phys. **82**, 1166 (1997)
- 32.312 J. Pastrnak, S. Pacesova, L. Roskovcova: Czech. J. Phys. **B24**, 1149 (1974)
- 32.313 F. Karel, J. Pastrnak, J. Hejduk, V. Losik: Phys. Stat. Sol. **15**, 693 (1966)
- 32.314 F. Karel, J. Mares: Czech. J. Phys. B **22**, 847 (1972)
- 32.315 G. E. Archangelskii, F. Karel, J. Mares, S. Pacesova, J. Pastrnak: Phys. Stat. Sol. **69**, 173 (1982)
- 32.316 R. G. Wilson, R. N. Schwartz, C. R. Abernathy, S. J. Pearton, N. Newman, M. Rubin, T. Fu, J. M. Zavada: Appl. Phys. Lett. **65**, 992 (1994)
- 32.317 J. D. MacKenzie, C. R. Abernathy, S. J. Pearton, U. Hömmerich, X. Wu, R. N. Schwartz, R. G. Wilson, J. M. Zavada: Appl. Phys. Lett. **69**, 2083 (1996)
- 32.318 X. Wu, U. Hömmerich, J. D. MacKenzie, C. R. Abernathy, S. J. Pearton, R. G. Wilson, R. N. Schwartz, J. M. Zavada: J. Lumin. **72–74**, 284 (1997)
- 32.319 S. J. Pearton, J. D. MacKenzie, C. R. Abernathy, U. Hömmerich, X. Wu, R. G. Wilson, R. N. Schwartz, J. M. Zavada, F. Ren: Appl. Phys. Lett. **71**, 1807 (1997)
- 32.320 J. Li, K. B. Nam, M. L. Nakarmi, J. Y. Lin, H. X. Jiang: Appl. Phys. Lett. **81**, 3365 (2002)
- 32.321 E. Kuokstis, J. Zhang, Q. Fareed, J. W. Yang, G. Simin, M. A. Khan, R. Gaska, M. Shur, C. Rojo, L. Schowalter: Appl. Phys. Lett. **81**, 2755 (2002)
- 32.322 K. B. Nam, J. Li, M. L. Nakarmi, J. Y. Lin, H. X. Jiang: Appl. Phys. Lett. **82**, 1694 (2003)
- 32.323 T. Onuma, S. F. Chichibu, T. Sota, K. Asai, S. Sumiya, T. Shibata, M. Tanaka: Appl. Phys. Lett. **81**, 652 (2002)
- 32.324 M. Morita, K. Tsubouchi, N. Mikoshiba: Jpn. J. Appl. Phys. **21**, 1102 (1982)
- 32.325 F. R. B. Hossain, X. Tang, K. Wongchotigul, M. G. Spencer: Proc. SPIE **42**, 2877 (1996)
- 32.326 X. Tang, F. R. B. Hossain, K. Wongchotigul, M. G. Spencer: Appl. Phys. Lett. **72**, 1501 (1998)
- 32.327 Y. Shishkin, R. P. Devaty, W. J. Choyke, F. Yun, T. King, H. Morkoç: Phys. Stat. Sol. (a) **188**, 591 (2001)
- 32.328 Q. Guo, A. Yoshida: Jpn. J. Appl. Phys. **33**, 2453 (1994)
- 32.329 P. B. Perry, R. F. Rutz: Appl. Phys. Lett. **33**, 319 (1978)
- 32.330 J. Pasternak, L. Roskovcova: Phys. Stat. Sol. **26**, 591 (1968)
- 32.331 T. Yodo, H. Yona, H. Ando, D. Nosei, Y. Harada: Appl. Phys. Lett. **80**, 968 (2002)
- 32.332 J. Bauer, L. Biste, D. Bolze: Phys. Stat. Sol. **39**, 173 (1977)
- 32.333 R. G. Gordon, D. M. Hoffmann, U. Riaz: J. Mater. Res. **6**, 5 (1991)
- 32.334 H. Demiryont, L. R. Thompson, G. J. Collins: Appl. Opt. **25**, 1311 (1986)
- 32.335 W. J. Meng, J. A. Sell, G. L. Eesley: J. Appl. Phys. **74**, 2411 (1993)
- 32.336 I. Akasaki, M. Hashimoto: Solid State Commun. **5**, 851 (1967)
- 32.337 A. T. Collins, E. C. Lightowlers, P. J. Dean: Phys. Rev. **158**, 833 (1967)

- 32.338 A. J. Noreika, M. H. Francombe, S. A. Zeitman: J. Vac. Sci. Technol. **6**, 194 (1969)
- 32.339 K. Osamura, N. Nakajima, Y. Murakami, P. H. Shingu, A. Ohtsuki: Solid State Commun. **46**, 3432 (1975)
- 32.340 T. Inushima, T. Yaguchi, A. Nagase, T. Shiraishi: Ins. Phys. Conf. Ser. **142**, 971 (1996)
- 32.341 A. Wakahara, T. Tsuchiya, A. Yoshida: J. Cryst. Growth **99**, 385 (1990)
- 32.342 K. Kubota, Y. Kobayashi, K. Fujimoto: J. Appl. Phys. **66**, 2984 (1989)
- 32.343 T. L. Tansley, C. P. Foley: J. Appl. Phys. **59**, 3241 (1986)
- 32.344 S. Yamaguchi, M. Kariya, S. Nitta, T. Takeuchi, C. Wetzel, H. Amano, I. Akasaki: Appl. Phys. Lett. **76**, 876 (2000)
- 32.345 V. Yu. Davydov, A. A. Klochikhin, R. P. Seisyan, V. V. Emtsev, S. V. Ivanov, F. Bechstedt, J. Furthmüller, H. Harima, A. V. Mudryi, J. Aderhold, O. Semchinova, J. Graul: Phys. Stat. Solidi (b) **229**, 1 (2002)
- 32.346 T. Inushima, V. V. Mamutin, V. A. Vekshin, S. V. Ivanov, T. Sakon, M. Motokawa, S. Ohoya: J. Crystal Growth **227–228**, 481 (2001)
- 32.347 J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager III, E. E. Haller, H. Lu, W. J. Schaff: Appl. Phys. Lett. **80**, 4741 (2002)
- 32.348 J. Wu, W. Walukiewicz, W. Shan, K. M. Yu, J. W. Ager III, E. E. Haller, H. Lu, W. J. Schaff: Phys. Rev. B **60**, 201403 (2002)
- 32.349 J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager III, E. E. Haller, H. Lu, W. J. Schaff, Y. Saito, Y. Nanishi: Appl. Phys. Lett. **80**, 3967 (2002)
- 32.350 K. Osamura, S. Naka, Y. Murakami: J. Appl. Phys. **46**, 3432 (1975)
- 32.351 A. Wakahara, T. Tsuchida, A. Yoshida: Vacuum **41**, 1071 (1990)
- 32.352 K. L. Westra, M. J. Brett: Thin Solid Films **192**, 234 (1990)
- 32.353 J. W. Trainor, K. Rose: J. Electron. Mater. **3**, 821 (1974)
- 32.354 Q. Guo, O. Kato, M. Fujisawa, A. Yoshida: Solid State Commun. **83**, 721 (1992)
- 32.355 Q. Guo, H. Ogawa, A. Yoshida: J. Electron. Spectrosc. Relat. Phenom. **79**, 9 (1996)
- 32.356 V. A. Tyagai, O. V. Snitko, A. M. Evstigneev, A. N. Krasiko: Phys. Stat. Sol. (b) **103**, 589 (1981)
- 32.357 F. Yun, M. A. Reshchikov, L. He, T. King, H. Morkoç, S. W. Novak, L. Wei: J. Appl. Phys. Rapid Commun. **92**, 4837 (2002)
- 32.358 S. Yoshida, S. Misawa, S. Gonda: J. Appl. Phys. **53**, 6844 (1982)
- 32.359 S. A. Nikishin, N. N. Faleev, A. S. Zubrilov, V. G. Antipov, H. Temkin: Appl. Phys. Lett. **76**, 3028 (2000)
- 32.360 W. Shan, J. W. Ager III, K. M. Yu, W. Walukiewicz, E. E. Haller, M. C. Martin, W. R. McKinney, W. Yang: J. Appl. Phys. **85**, 8505 (1999)
- 32.361 Ü. Özgür, G. Webb-Wood, H. O. Everitt, F. Yun, H. Morkoç: Appl. Phys. Lett. **79**, 4103 (2001)
- 32.362 J. Wagner, H. Obloh, M. Kunzer, M. Maier, K. Kohler, B. Johs: J. Appl. Phys. **89**, 2779 (2000)
- 32.363 H. Jiang, G. Y. Zhao, H. Ishikawa, T. Egawa, T. Jimbo, M. J. Umeno: Appl. Phys. **89**, 1046 (2001)
- 32.364 T. J. Ochaliski, B. Gil, P. Lefebvre, M. Grandjean, M. Leroux, J. Massies, S. Nakamura, H. Morkoç: Appl. Phys. Lett. **74**, 3353 (1999)
- 32.365 M. A. Khan, J. M. Van Hove, J. N. Kuznia, D. T. Olson: Appl. Phys. Lett. **58**, 2408 (1991)
- 32.366 T. Tanaka, A. Watanabe, H. Amano, Y. Kobayashi, I. Akasaki, S. Yamazaki, M. Koike: Appl. Phys. Lett. **65**, 593 (1994)
- 32.367 M. Stutzmann, O. Ambacher, A. Cros, M. S. Brandt, H. Angerer, R. Dimitrov, N. Reinacher, T. Metzger, R. Hopler, D. Brunner, F. Freudenberger, R. Handschuh, Ch. Deger: presented at the E-MRS Straßburg, Symposium L (1997)
- 32.368 S. Nakamura, T. Mukai: J. Vac. Sci. Technol. A **13**, 6844 (1995)
- 32.369 H. Amano, T. Takeuchi, S. Sota, H. Sakai, I. Akasaki: In: *III–V nitrides*, Vol. 449, ed. by F. A. Ponce, T. D. Moustakas, I. Akasaki, B. Menemar (MRS Proc., Pittsburgh, Pennsylvania 1997) p. 1143
- 32.370 K. P. O'Donnell, R. W. Martin, C. Trager-Cowan, M. E. White, K. Esona, C. Deatcher, P. G. Middleton, K. Jacobs, W. van der Stricht, C. Merlet, B. Gil, A. Vantomme, J. F. W. Mosselmans: Mater. Sci. Eng. B **82**, 194 (2001)
- 32.371 T. Nagatomo, T. Kuboyama, H. Minamino, O. Omoto: Jpn. J. Appl. Phys. **28**, L1334 (1989)
- 32.372 W. Shan, W. Walukiewicz, E. E. Haller, B. D. Little, J. J. Song, M. D. McCluskey, N. M. Johnson, Z. C. Feng, M. Schurman, R. A. Stall: J. Appl. Phys. **84**, 4452 (1998)
- 32.373 S. Pereira, M. R. Correia, T. Monteiro, E. Pereira, E. Alves, A. D. Sequeira, N. Franco: Appl. Phys. Lett. **78**, 2137 (2001)
- 32.374 I. Akasaki, H. Amano: Jpn. J. Appl. Phys. **36**, 5393 (1997)
- 32.375 S. Yamasaki, S. Asami, N. Shibata, M. Koike, K. Manabe, T. Tanaka, H. Amano, I. Akasaki: Appl. Phys. Lett. **66**, 1112 (1995)
- 32.376 N. Yoshimoto, T. Matsuoka, A. Katsui: Appl. Phys. Lett. **59**, 2251 (1991)
- 32.377 S. Nakamura, T. Mukai, M. Seno: Jpn. J. Appl. Phys. **31**, L16 (1993)
- 32.378 S. Nakamura, N. Iwasa, S. Nagahama: Jpn. J. Appl. Phys. **32**, L338 (1993)
- 32.379 K. S. Kim, A. Saxler, P. Kung, R. Razeghi, K. Y. Lim: Appl. Phys. Lett. **71**, 800 (1997)
- 32.380 K. Starosta: Phys. Status Solidi A **68**, K55 (1981)
- 32.381 T. J. Kistenmacher, S. A. Ecelberger, W. A. Bryden: J. Appl. Phys. **74**, 1684 (1993)
- 32.382 S. Yamaguchi, M. Kariya, S. Nitta, T. Takeuchi, C. Wetzel, H. Amano, I. Akasaki: Appl. Phys. Lett. **73**, 830 (1998)
- 32.383 G. Davies: Phys. Rep. **176**, 83 (1989)

- 32.384 W. R. Bryden, T. J. Kistenmacher: Electrical transport properties of InN, GaInN and AlInN. In: *Properties of Group III Nitrides*, ed. by J. H. Edgar (INSPEC, London 1994)
- 32.385 S. M. Bedair, F. G. McIntosh, J. C. Roberts, E. L. Piner, K. S. Boutros, N. A. El-Masry: *J. Crystal Growth* **178**, 32 (1997)
- 32.386 S. N. Mohammad, A. Salvador, H. Morkoç: *Proc. IEEE* **83**, 1306 (1995)
- 32.387 M.-Y. Ryu, C. Q. Chen, E. Kuokstis, J. W. Yang, G. Simin, M. A. Khan: *Appl. Phys. Lett.* **80**, 3730 (2002)
- 32.388 K. Kim, W. R. L. Lambrecht, B. Segall: *Phys. Rev. B* **53**, 7018 (1997)

33. Electron Transport Within the III–V Nitride Semiconductors, GaN, AlN, and InN: A Monte Carlo Analysis

The III–V nitride semiconductors, gallium nitride, aluminium nitride, and indium nitride, have been recognized as promising materials for novel electronic and optoelectronic device applications for some time now. Since informed device design requires a firm grasp of the material properties of the underlying electronic materials, the electron transport that occurs within these III–V nitride semiconductors has been the focus of considerable study over the years. In an effort to provide some perspective on this rapidly evolving field, in this paper we review analyses of the electron transport within these III–V nitride semiconductors. In particular, we discuss the evolution of the field, compare and contrast results obtained by different researchers, and survey the current literature. In order to narrow the scope of this chapter, we will primarily focus on electron transport within bulk wurtzite gallium nitride, aluminium nitride, and indium nitride for this analysis. Most of our discussion will focus on results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these materials, our results conforming with state-of-the-art III–V nitride semiconductor orthodoxy. Steady-state and transient electron transport results are presented. We conclude our discussion by presenting some recent developments on the electron transport within these materials.

33.1 Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach	806
33.1.1 The Boltzmann Transport Equation	807
33.1.2 Our Ensemble Semi-Classical Monte Carlo Simulation Approach	808
33.1.3 Parameter Selections for Bulk Wurtzite GaN, AlN, and InN	808
33.2 Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN	810
33.2.1 Steady-State Electron Transport Within Bulk Wurtzite GaN	811
33.2.2 Steady-State Electron Transport: A Comparison of the III–V Nitride Semiconductors with GaAs	812
33.2.3 Influence of Temperature on the Electron Drift Velocities Within GaN and GaAs	812
33.2.4 Influence of Doping on the Electron Drift Velocities Within GaN and GaAs	815
33.2.5 Electron Transport in AlN	816
33.2.6 Electron Transport in InN	818
33.2.7 Transient Electron Transport	820
33.2.8 Electron Transport: Conclusions	822
33.3 Electron Transport Within III–V Nitride Semiconductors: A Review	822
33.3.1 Evolution of the Field	822
33.3.2 Recent Developments	824
33.3.3 Future Perspectives	825
33.4 Conclusions	826
References	826

The III–V nitride semiconductors, gallium nitride (GaN), aluminium nitride (AlN), and indium nitride (InN), have been known as promising materials for novel electronic and optoelectronic device applications for some time now [33.1–4]. In terms of electronics, their wide energy gaps, large breakdown fields, high thermal conductivities, and favorable electron transport characteristics, make GaN, AlN, and InN, and alloys of these materials, ideally suited for novel

high-power and high-frequency electron device applications. On the optoelectronics front, the direct nature of the energy gaps associated with GaN, AlN, and InN, make this family of materials, and its alloys, well suited for novel optoelectronic device applications in the visible and ultraviolet frequency range. While initial efforts to study these materials were hindered by growth difficulties, recent improvements in material quality have made the realization of a number of

III–V nitride semiconductor-based electronic [33.5–9] and optoelectronic [33.9–12] devices possible. These developments have fueled considerable interest in these III–V nitride semiconductors.

In order to analyze and improve the design of III–V nitride semiconductor-based devices, an understanding of the electron transport that occurs within these materials is necessary. Electron transport within bulk GaN, AlN, and InN has been examined extensively over the years [33.13–32]. Unfortunately, uncertainty in the material parameters associated with GaN, AlN, and InN remains a key source of ambiguity in the analysis of the electron transport within these materials [33.32]. In addition, some recent experimental [33.33] and theoretical [33.34] developments have cast doubt upon the validity of widely accepted notions upon which our understanding of the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, has evolved. Another confounding matter is the sheer volume of research activity being performed on the electron transport within these materials, presenting the researcher with a dizzying array of seemingly disparate approaches and results. Clearly, at this critical juncture at least, our understanding of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, remains in a state of flux.

In order to provide some perspective on this rapidly evolving field, we aim to review analyses of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, within this paper. In particular, we will discuss the evolution of the field and survey the current literature. In order to narrow the scope of this review, we will primarily focus on the electron transport within bulk wurtzite GaN, AlN, and InN for the purposes of this paper. Most of our discussion will focus upon results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within

these materials, our results conforming with state-of-the-art III–V nitride semiconductor orthodoxy. We hope that researchers in the field will find this review useful and informative.

We begin our review with the Boltzmann transport equation, which underlies most analyses of the electron transport within semiconductors. The ensemble semi-classical three-valley Monte Carlo simulation approach that we employ in order to solve this Boltzmann transport equation is then discussed. The material parameters corresponding to bulk wurtzite GaN, AlN, and InN are then presented. We then use these material parameter selections and our ensemble semi-classical three-valley Monte Carlo simulation approach to determine the nature of the steady-state and transient electron transport within the III–V nitride semiconductors. Finally, we present some recent developments on the electron transport within these materials.

This paper is organized in the following manner. In Sect. 33.1, we present the Boltzmann transport equation and our ensemble semi-classical three-valley Monte Carlo simulation approach that we employ in order to solve this equation for the III–V nitride semiconductors, GaN, AlN, and InN. The material parameters, corresponding to bulk wurtzite GaN, AlN, and InN, are also presented in Sect. 33.1. Then, in Sect. 33.2, using results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these III–V nitride semiconductors, we study the nature of the steady-state electron transport that occurs within these materials. Transient electron transport within the III–V nitride semiconductors is also discussed in Sect. 33.2. A review of the III–V nitride semiconductor electron transport literature, in which the evolution of the field is discussed and a survey of the current literature is presented, is then featured in Sect. 33.3. Finally, conclusions are provided in Sect. 33.4.

33.1 Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach

The electrons within a semiconductor are in a perpetual state of motion. In the absence of an applied electric field, this motion arises as a result of the thermal energy that is present, and is referred to as thermal motion. From the perspective of an individual electron, thermal motion may be viewed as a series of trajectories, interrupted by a series of random scattering events. Scattering may arise

as a result of interactions with the lattice atoms, impurities, other electrons, and defects. As these interactions lead to electron trajectories in all possible directions, i.e., there is no preferred direction, while individual electrons will move from one location to another, when taken as an ensemble, and assuming that the electrons are in thermal equilibrium, the overall electron distribu-

tion will remain static. Accordingly, no net current flow occurs.

With the application of an applied electric field, E , each electron in the ensemble will experience a force, $-qE$. While this force may have a negligible impact upon the motion of any given individual electron, taken as an ensemble, the application of such a force will lead to a net aggregate motion of the electron distribution. Accordingly, a net current flow will occur, and the overall electron ensemble will no longer be in thermal equilibrium. This movement of the electron ensemble in response to an applied electric field, in essence, represents the fundamental issue at stake when we study the electron transport within a semiconductor.

In this section, we provide a brief tutorial on the issues at stake in our analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. We begin our analysis with an introduction to the Boltzmann transport equation. This equation describes how the electron distribution function evolves under the action of an applied electric field, and underlies the electron transport within bulk semiconductors. We then introduce the Monte Carlo simulation approach to solving this Boltzmann transport equation, focusing on the ensemble semi-classical three-valley Monte Carlo simulation approach used in our simulations of the electron transport within the III–V nitride semiconductors. Finally, we present the material parameters corresponding to bulk wurtzite GaN, AlN, and InN.

This section is organized in the following manner. In Sect. 33.1.1, the Boltzmann transport equation is introduced. Then, in Sect. 33.1.2, our ensemble semi-classical three-valley Monte Carlo simulation approach to solving this Boltzmann transport equation is presented. Finally, in Sect. 33.1.3, our material parameter selections, corresponding to bulk wurtzite GaN, AlN, and InN, are presented.

33.1.1 The Boltzmann Transport Equation

An electron ensemble may be characterized by its distribution function, $f(\mathbf{r}, \mathbf{p}, t)$, where \mathbf{r} denotes the position, \mathbf{p} represents the momentum, and t indicates time. The response of this distribution function to an applied electric field, E , is the issue at stake when one investigates the electron transport within a semiconductor. When the dimensions of the semiconductor are large, and quantum effects are negligible, the ensemble of electrons may be treated as a continuum, so the corpuscular nature of the individual electrons within the ensemble, and the attendant complications which

arise, may be neglected. In such a circumstance, the evolution of the distribution function, $f(\mathbf{r}, \mathbf{p}, t)$, may be determined using the Boltzmann transport equation. In contrast, when the dimensions of the semiconductor are small, and quantum effects are significant, then the Boltzmann transport equation, and its continuum description of the electron ensemble, is no longer valid. In such a case, it is necessary to adopt quantum transport methods in order to study the electron transport within the semiconductor [33.35].

For the purposes of this analysis, we will focus on the electron transport within bulk semiconductors, i.e., semiconductors of sufficient dimensions so that the Boltzmann transport equation is valid. *Ashcroft* and *Mermin* [33.36] demonstrated that this equation may be expressed as

$$\frac{\partial f}{\partial t} = -\dot{\mathbf{p}} \cdot \nabla_{\mathbf{p}} f - \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f + \left. \frac{\partial f}{\partial t} \right|_{\text{scat}}. \quad (33.1)$$

The first term on the right-hand side of (33.1) represents the change in the distribution function due to external forces applied to the system. The second term on the right-hand side of (33.1) accounts for the electron diffusion which occurs. The final term on the right-hand side of (33.1) describes the effects of scattering.

Owing to its fundamental importance in the analysis of the electron transport within semiconductors, a number of techniques have been developed over the years in order to solve the Boltzmann transport equation. Approximate solutions to the Boltzmann transport equation, such as the displaced Maxwellian distribution function approach of *Ferry* [33.14] and *Das* and *Ferry* [33.15] and the nonstationary charge transport analysis of *Sandborn* et al. [33.37], have proven useful. Low-field approximate solutions have also proven elementary and insightful [33.17, 20, 38]. A number of these techniques have been applied to the analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN [33.14, 15, 17, 20, 38, 39]. Alternatively, more sophisticated techniques have been developed which solve the Boltzmann transport equation directly. These techniques, while allowing for a rigorous solution of the Boltzmann transport equation, are rather involved, and require intense numerical analysis. They are further discussed by *Nag* [33.40].

For studies of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, by far the most common approach to solving the Boltzmann transport equation has been the ensemble semi-classical Monte Carlo simulation approach. Of the III–V nitride semiconductors, the electron transport within GaN

has been studied the most extensively using this ensemble Monte Carlo simulation approach [33.13, 16, 18, 19, 21, 22, 27, 29, 32], with AlN [33.24, 25, 29] and InN [33.23, 28, 29, 31] less so. The Monte Carlo simulation approach has also been used to study the electron transport within the two-dimensional electron gas of the AlGaIn/GaN interface which occurs in high electron mobility AlGaIn/GaN field-effect transistors [33.41, 42].

At this point, it should be noted that the complete solution of the Boltzmann transport equation requires the resolution of both steady-state and transient responses. Steady-state electron transport refers to the electron transport that occurs long after the application of an applied electric field, i. e., once the electron ensemble has settled to a new equilibrium state (we are not necessarily referring to thermal equilibrium here, since thermal equilibrium is only achieved in the absence of an applied electric field). As the distribution function is difficult to visualize quantitatively, researchers typically study the dependence of the electron drift velocity (the average electron velocity determined by statistically averaging over the entire electron ensemble) on the applied electric field in the analysis of steady-state electron transport; in other words, they determine the velocity–field characteristic. Transient electron transport, by way of contrast, refers to the transport that occurs while the electron ensemble is evolving into its new equilibrium state. Typically, it is characterized by studying the dependence of the electron drift velocity on the time elapsed, or the distance displaced, since the electric field was initially applied. Both steady-state and transient electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, are reviewed within this paper.

33.1.2 Our Ensemble Semi-Classical Monte Carlo Simulation Approach

For the purposes of our analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, we employ ensemble semi-classical Monte Carlo simulations. A three-valley model for the conduction band is employed. Nonparabolicity is considered in the lowest conduction band valley, this nonparabolicity being treated through the application of the Kane model [33.43].

In the Kane model, the energy band of the Γ valley is assumed to be nonparabolic, spherical, and of the form

$$\frac{\hbar^2 k^2}{2m^*} = E(1 + \alpha E), \quad (33.2)$$

where $\hbar k$ denotes the crystal momentum, E represents the energy above the minimum, m^* is the effective mass, and the nonparabolicity coefficient, α , is given by

$$\alpha = \frac{1}{E_g} \left(1 - \frac{m^*}{m_e}\right)^2, \quad (33.3)$$

where m_e and E_g denote the free electron mass and the energy gap, respectively [33.43].

The scattering mechanisms considered in our analysis are (1) ionized impurity, (2) polar optical phonon, (3) piezoelectric [33.44, 45], and (4) acoustic deformation potential. Intervalley scattering is also considered. Piezoelectric scattering is treated using the well established zinc blende scattering rates, and so a suitably transformed piezoelectric constant, e_{14} , must be selected. This may be achieved through the transformation suggested by Bykhovski et al. [33.44, 45]. We also assume that all donors are ionized and that the free electron concentration is equal to the dopant concentration. The motion of three thousand electrons is examined in our steady-state electron transport simulations, while the motion of ten thousand electrons is considered in our transient electron transport simulations. The crystal temperature is set to 300 K and the doping concentration is set to 10^{17} cm^{-3} in all cases, unless otherwise specified. Electron degeneracy effects are accounted for by means of the rejection technique of Lugli and Ferry [33.46]. Electron screening is also accounted for following the Brooks–Herring method [33.47]. Further details of our approach are discussed in the literature [33.16, 21–24, 29, 32, 48].

33.1.3 Parameter Selections for Bulk Wurtzite GaN, AlN, and InN

The material parameter selections, used for our simulations of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, are tabulated in Table 33.1. These parameter selections are the same as those employed by Foutz et al. [33.29]. While the band structures corresponding to bulk wurtzite GaN, AlN, and InN are still not agreed upon, the band structures of Lambrecht and Segall [33.49] are adopted for the purposes of this analysis. For the case of bulk wurtzite GaN, the analysis of Lambrecht and Segall [33.49] suggests that the lowest point in the conduction band is located at the center of the Brillouin zone, at the Γ point, the first upper conduction band valley minimum also occurring at the Γ point, 1.9 eV above the lowest point in the conduction band, the second upper conduction

Table 33.1 The material parameter selections corresponding to bulk wurtzite GaN, AlN, and InN. These parameter selections are from Foutz et al. [33.29]

Parameter	GaN	AlN	InN
Mass density (g/cm ³)	6.15	3.23	6.81
Longitudinal sound velocity (cm/s)	6.56×10^5	9.06×10^5	6.24×10^5
Transverse sound velocity (cm/s)	2.68×10^5	3.70×10^5	2.55×10^5
Acoustic deformation potential (eV)	8.3	9.5	7.1
Static dielectric constant	8.9	8.5	15.3
High-frequency dielectric constant	5.35	4.77	8.4
Effective mass (Γ_1 valley)	$0.20 m_e$	$0.48 m_e$	$0.11 m_e$
Piezoelectric constant, e_{14} (C/cm ²)	3.75×10^{-5}	9.2×10^{-5}	3.75×10^{-5}
Direct energy gap (eV)	3.39	6.2	1.89
Optical phonon energy (meV)	91.2	99.2	89.0
Intervalley deformation potentials (eV/cm)	10^9	10^9	10^9
Intervalley phonon energies (meV)	91.2	99.2	89.0

Table 33.2 The valley parameter selections corresponding to bulk wurtzite GaN, AlN, and InN. These parameter selections are from Foutz et al. [33.29]. These parameters were originally determined from the band structural calculations of Lambrecht and Segall [33.49]

	Valley number	1	2	3
GaN	Valley location	Γ_1	Γ_2	L–M
	Valley degeneracy	1	1	6
	Effective mass	$0.2 m_e$	m_e	m_e
	Intervalley energy separation (eV)	—	1.9	2.1
	Energy gap (eV)	3.39	5.29	5.49
	Nonparabolicity (eV ^{−1})	0.189	0.0	0.0
AlN	Valley location	Γ_1	L–M	K
	Valley degeneracy	1	6	2
	Effective mass	$0.48 m_e$	m_e	m_e
	Intervalley energy separation (eV)	—	0.7	1.0
	Energy gap (eV)	6.2	6.9	7.2
	Nonparabolicity (eV ^{−1})	0.044	0.0	0.0
InN	Valley location	Γ_1	A	Γ_2
	Valley degeneracy	1	1	1
	Effective mass	$0.11 m_e$	m_e	m_e
	Intervalley energy separation (eV)	—	2.2	2.6
	Energy gap (eV)	1.89	4.09	4.49
	Nonparabolicity (eV ^{−1})	0.419	0.0	0.0

band valley minima occurring along the symmetry lines between the L and M points, 2.1 eV above the lowest point in the conduction band; see Table 33.2. For the case of bulk wurtzite AlN, the analysis of Lambrecht and Segall [33.49] suggests that the lowest point in the conduction band is located at the center of the Brillouin zone, at the Γ point, the first upper conduction band valley minima occurring along the symmetry lines between the L and M points, 0.7 eV above the lowest

point in the conduction band, the second upper conduction band valley minima occurring at the K points, 1 eV above the lowest point in the conduction band; see Table 33.2. For the case of bulk wurtzite InN, the analysis of Lambrecht and Segall [33.49] suggests that the lowest point in the conduction band is located at the center of the Brillouin zone, at the Γ point, the first upper conduction band valley minimum occurring at the A point, 2.2 eV above the lowest point in the conduction band,

the second upper conduction band valley minimum occurring at the Γ point, 2.6 eV above the lowest point in the conduction band; see Table 33.2. We ascribe an effective mass equal to the free electron mass, m_e , to all of the upper conduction band valleys. The nonparabolicity coefficient, α , corresponding to each upper conduction band valley is set to zero, so the upper conduction band valleys are assumed to be completely parabolic. For our simulations of the electron transport within gallium arsenide (GaAs), the material parameters employed are mostly from *Littlejohn* et al. [33.50], although it should be noted that the mass density, the energy gap, and the sound velocities are from *Blakemore* [33.51].

It should be noted that the energy gap associated with InN has been the subject of some controversy since 2002. The pioneering experimental results of *Tansley* and *Foley* [33.52], reported in 1986, suggested that InN has an energy gap of 1.89 eV. This value has been used extensively in Monte Carlo simulations of the electron transport within this material since that time [33.23, 28, 29, 31]; typically, the influence of the energy gap on the electron transport occurs through its impact on the nonparabolicity coefficient, α . In 2002, *Davydov* et al. [33.53], *Wu* et al. [33.54], and *Matsuoka* et al. [33.55], presented experimental evidence which instead suggests a considerably smaller energy gap for InN, around 0.7 eV. As this new result is still the subject of some controversy, we adopt the traditional *Tansley* and *Foley* [33.52] energy gap value for the purposes of

our present analysis, noting that even if the newer value for the energy gap was adopted, it would only change our electron transport results marginally; the sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the nonparabolicity coefficient, α , has been explored, in detail, by *O’Leary* et al. [33.32].

The band structure associated with bulk wurtzite GaN has also been the focus of some controversy. In particular, *Brazel* et al. [33.56] employed ballistic electron emission microscopy measurements in order to demonstrate that the first upper conduction band valley occurs only 340 meV above the lowest point in the conduction band for this material. This contrasts rather dramatically with more traditional results, such as the calculation of *Lambrecht* and *Segall* [33.49], which instead suggest that the first upper conduction band valley minimum within wurtzite GaN occurs about 2 eV above the lowest point in the conduction band. Clearly, this will have a significant impact upon the results. While the results of *Brazel* et al. [33.56] were reported in 1997, electron transport simulations adopted the more traditional intervalley energy separation of about 2 eV until relatively recently. Accordingly, we have adopted the more traditional intervalley energy separation for the purposes of our present analysis. The sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the intervalley energy separation has been explored, in detail, by *O’Leary* et al. [33.32].

33.2 Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN

The current interest in the III–V nitride semiconductors, GaN, AlN, and InN, is primarily being fueled by the tremendous potential of these materials for novel electronic and optoelectronic device applications. With the recognition that informed electronic and optoelectronic device design requires a firm understanding of the nature of the electron transport within these materials, electron transport within the III–V nitride semiconductors has been the focus of intensive investigation over the years. The literature abounds with studies on steady-state and transient electron transport within these materials [33.13–34, 38, 39, 41, 42, 48]. As a result of this intense flurry of research activity, novel III–V nitride semiconductor-based devices are starting to be deployed in today’s commercial products. Future developments in the III–V nitride semiconductor field will undoubtedly require an even deeper understand-

ing of the electron transport mechanisms within these materials.

In the previous section, we presented details of the Monte Carlo simulation approach that we employ for the analysis of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. In this section, an overview of the steady-state and transient electron transport results we obtained from these Monte Carlo simulations is provided. In the first part of this section, we focus upon bulk wurtzite GaN. In particular, the velocity–field characteristic associated with this material will be examined in detail. Then, an overview of our steady-state electron transport results, corresponding to the three III–V nitride semiconductors under consideration in this analysis, will be given, and a comparison with the more conventional III–V compound semiconductor, GaAs, will be presented. A comparison between the tem-

perature dependence of the velocity–field characteristics associated with GaN and GaAs will then be presented, and our Monte Carlo results will be used to account for the differences in behavior. A similar analysis will be presented for the doping dependence. Next, detailed simulation results for AlN and InN will be presented. Finally, the transient electron transport that occurs within the III–V nitride semiconductors, GaN, AlN, and InN, is determined and compared with that in GaAs.

This section is organized in the following manner. In Sect. 33.2.1, the velocity–field characteristic associated with bulk wurtzite GaN is presented and analyzed. Then, in Sect. 33.2.2, the velocity–field characteristics associated with the III–V nitride semiconductors under consideration in this analysis will be compared and contrasted with that of GaAs. The sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the crystal temperature will then be examined in Sect. 33.2.3, and a comparison with that corresponding to GaAs presented. In Sect. 33.2.4, the sensitivity of the velocity–field characteristic associated with bulk wurtzite GaN to variations in the doping concentration level will be explored, and a comparison with that corresponding to GaAs presented. The velocity–field characteristics associated with AlN and InN will then be examined in Sect. 33.2.5 and Sect. 33.2.6, respectively. Our transient electron transport analysis results are then presented in Sect. 33.2.7. Finally, the conclusions of this electron transport analysis are summarized in Sect. 33.2.8.

33.2.1 Steady-State Electron Transport Within Bulk Wurtzite GaN

Our examination of results begins with GaN, the most commonly studied III–V nitride semiconductor. The velocity–field characteristic associated with this material is presented in Fig. 33.1. This result was obtained through our Monte Carlo simulations of the electron transport within this material for the bulk wurtzite GaN parameter selections specified in Table 33.1 and Table 33.2; the crystal temperature was set to 300 K and the doping concentration to 10^{17} cm^{-3} . We see that for applied electric fields in excess of 140 kV/cm, the electron drift velocity decreases, eventually saturating at $1.4 \times 10^7 \text{ cm/s}$ for high applied electric fields. By examining the results of our Monte Carlo simulation further, an understanding of this result becomes clear.

First, we discuss the results at low applied electric fields, i. e., applied electric fields of less than 30 kV/cm. This is referred to as the linear regime of electron trans-

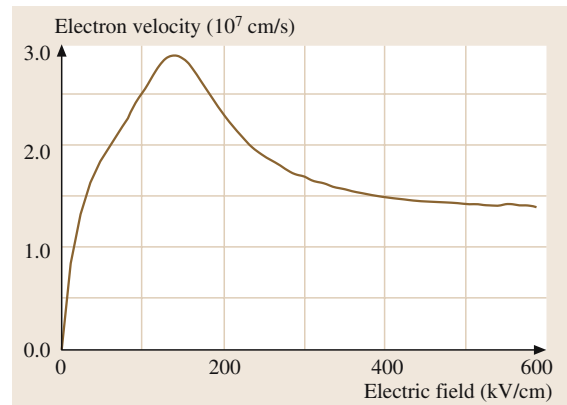


Fig. 33.1 The velocity–field characteristic associated with bulk wurtzite GaN. Like many other compound semiconductors, the electron drift velocity reaches a peak, and at higher applied electric fields it decreases until it saturates

port as the electron drift velocity is well characterized by the low-field electron drift mobility, μ , in this regime, i. e., a linear low-field electron drift velocity dependence on the applied electric field $v_d = \mu E$, applies in this regime. Examining the distribution function for this regime, we find that it is very similar to the zero-field distribution function with a slight shift in the direction opposite to the applied electric field. In this regime, the average electron energy remains relatively low, with most of the energy gained from the applied electric field being transferred into the lattice through polar optical phonon scattering.

If we examine the average electron energy as a function of the applied electric field, shown in Fig. 33.2, we see that there is a sudden increase at around 100 kV/cm. In order to understand why this increase occurs, we note that the dominant energy loss mechanism for many of the III–V compound semiconductors, including GaN, is polar optical phonon scattering. When the applied electric field is less than 100 kV/cm, all of the energy that the electrons gain from the applied electric field is lost through polar optical phonon scattering. The other scattering mechanisms, i. e., ionized impurity scattering, piezoelectric scattering and acoustic deformation potential scattering, do not remove energy from the electron ensemble: they are elastic scattering mechanisms. However, beyond a certain critical applied electric field strength, the polar optical phonon scattering mechanism can no longer remove all of the energy gained from the applied electric field. Other scattering mechanisms must start to play a role if the electron ensemble is to remain in equilibrium. The average electron energy increases

until intervalley scattering begins and an energy balance is re-established.

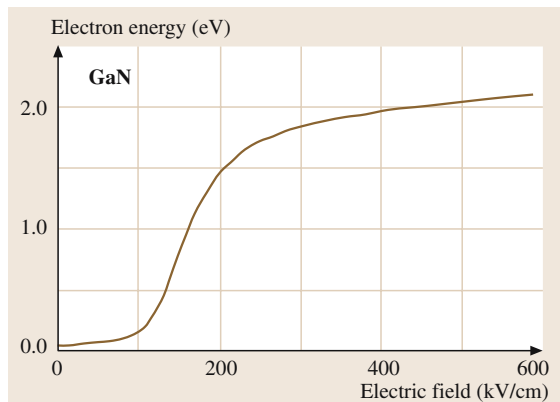


Fig. 33.2 The average electron energy as a function of the applied electric field for bulk wurtzite GaN. Initially, the average electron energy remains low, only slightly higher than the thermal energy, $\frac{3}{2}k_B T$. At 100 kV/cm, however, the average electron energy increases dramatically. This increase is due to the fact that the polar optical phonon scattering mechanism can no longer absorb all of the energy gained from the applied electric field

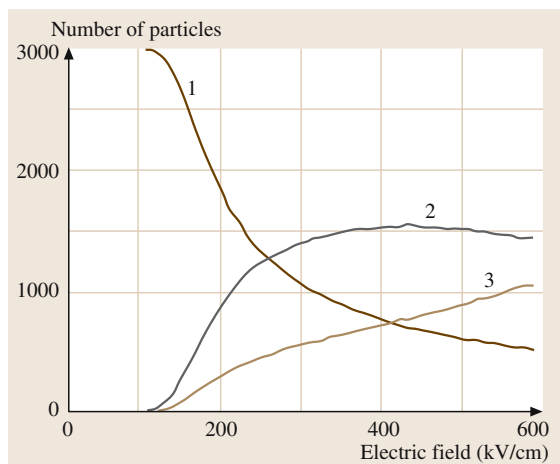


Fig. 33.3 The valley occupancy as a function of the applied electric field for the case of bulk wurtzite GaN. Soon after the average electron energy increases, electrons begin to transfer to the upper valleys of the conduction band. Three thousand electrons were employed for this simulation. The valleys are labeled 1, 2, and 3, in accordance with their energy minima; the lowest energy valley is valley 1, the next higher energy valley is valley 2, and the highest energy valley is valley 3

As the applied electric field is increased beyond 100 kV/cm, the average electron energy increases until a substantial fraction of the electrons have acquired enough energy in order to transfer into the upper valleys. As the effective mass of the electrons in the upper valleys is greater than that in the lowest valley, the electrons in the upper valleys will be slower. As more electrons transfer to the upper valleys (Fig. 33.3), the electron drift velocity decreases. This accounts for the negative differential mobility observed in the velocity–field characteristic depicted in Fig. 33.1.

Finally, at high applied electric fields, the number of electrons in each valley saturates. It can be shown that in the high-field limit the number of electrons in each valley is proportional to the product of the density of states of that particular valley and the corresponding valley degeneracy. At this point, the electron drift velocity stops decreasing and achieves saturation.

Thus far, electron transport results corresponding to bulk wurtzite GaN have been presented and discussed qualitatively. It should be noted, however, that the same phenomenon that occurs in the velocity–field characteristic associated with GaN also occurs for the other III–V nitride semiconductors, AlN and InN. The importance of polar optical phonon scattering when determining the nature of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, will become even more apparent later, as it will be used to account for much of the electron transport behavior within these materials.

33.2.2 Steady-State Electron Transport: A Comparison of the III–V Nitride Semiconductors with GaAs

Setting the crystal temperature to 300 K and the level of doping to 10^{17} cm^{-3} , the velocity–field characteristics associated with the III–V nitride semiconductors under consideration in this analysis – GaN, AlN, and InN – are contrasted with that of GaAs in Fig. 33.4. We see that each of these III–V compound semiconductors achieves a peak in its velocity–field characteristic. InN achieves the highest steady-state peak electron drift velocity, $4.1 \times 10^7 \text{ cm/s}$ at an applied electric field of 65 kV/cm. This contrasts with the case of GaN, $2.9 \times 10^7 \text{ cm/s}$ at 140 kV/cm, and that of AlN, $1.7 \times 10^7 \text{ cm/s}$ at 450 kV/cm. For GaAs, the peak electron drift velocity of $1.6 \times 10^7 \text{ cm/s}$ occurs at a much lower applied electric field than that for the III–V nitride semiconductors (only 4 kV/cm).

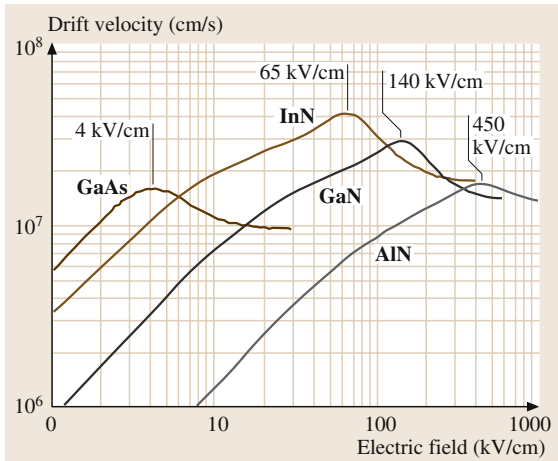


Fig. 33.4 A comparison of the velocity–field characteristics associated with the III–V nitride semiconductors, GaN, AlN, and InN, with that associated with GaAs. After [33.29] with permission, copyright AIP

33.2.3 Influence of Temperature on the Electron Drift Velocities Within GaN and GaAs

The temperature dependence of the velocity–field characteristic associated with bulk wurtzite GaN is now examined. Figure 33.5a shows how the velocity–field characteristic associated with bulk wurtzite GaN varies as the crystal temperature is increased from 100 to 700 K, in increments of 200 K. The upper limit, 700 K, is chosen as it is the highest operating temperature that

may be expected for AlGaN/GaN power devices. To highlight the difference between the III–V nitride semiconductors with more conventional III–V compound semiconductors, such as GaAs, Monte Carlo simulations of the electron transport within GaAs have also been performed under the same conditions as GaN. Figure 33.5b shows the results of these simulations. Note that the electron drift velocity for the case of GaN is much less sensitive to changes in temperature than that associated with GaAs.

To quantify this dependence further, the low-field electron drift mobility, the peak electron drift velocity, and the saturation electron drift velocity are plotted as a function of the crystal temperature in Fig. 33.6, these results being determined from our Monte Carlo simulations of the electron transport within these materials. For both GaN and GaAs, it is found that all of these electron transport metrics diminish as the crystal temperature is increased. As may be seen through an inspection of Fig. 33.5, the peak and saturation electron drift velocities do not drop as much in GaN as they do in GaAs in response to increases in the crystal temperature. The low-field electron drift mobility in GaN, however, is seen to fall quite rapidly with temperature, this drop being particularly severe for temperatures at and below room temperature. This property will likely have an impact on high-power device performance.

Delving deeper into our Monte Carlo results yields clues to the reason for this variation in temperature dependence. First, we examine the polar optical phonon scattering rate as a function of the applied electric field strength. Figure 33.7 shows that the scattering rate only

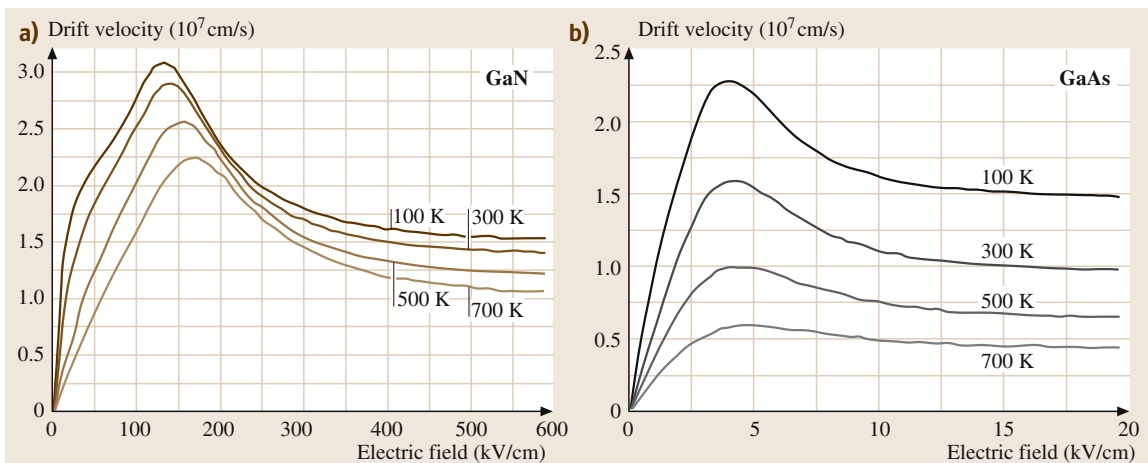


Fig. 33.5a,b A comparison of the temperature dependence of the velocity–field characteristics associated with (a) GaN and (b) GaAs. GaN maintains a higher electron drift velocity with increased temperatures than GaAs does

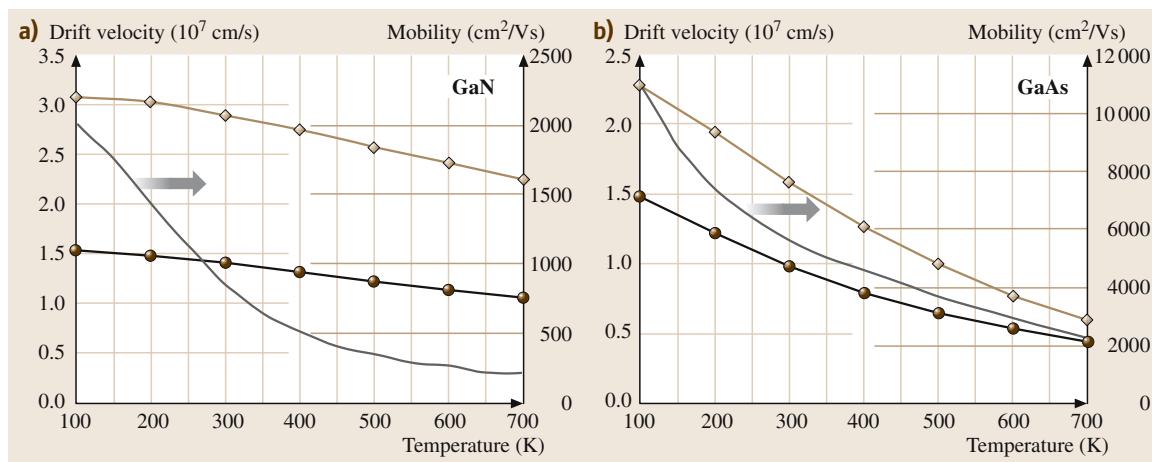


Fig. 33.6a,b A comparison of the temperature dependence of the low-field electron drift mobility (*solid lines*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*) for (a) GaN and (b) GaAs. The low-field electron drift mobility of GaN drops quickly with increasing temperature, but its peak and saturation electron drift velocities are less sensitive to increases in temperature than GaAs

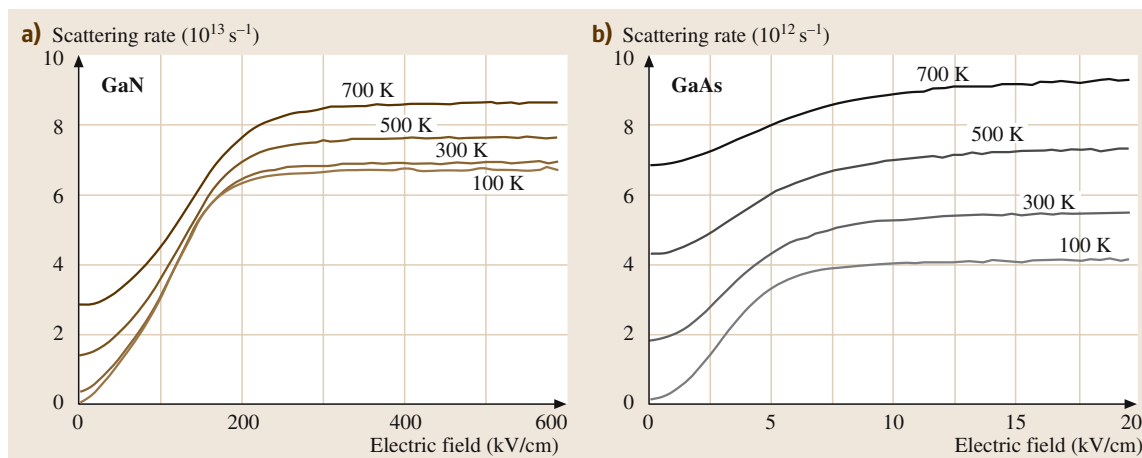


Fig. 33.7a,b A comparison of the polar optical phonon scattering rates as a function of the applied electric field strength for various crystal temperatures for (a) GaN and (b) GaAs. Polar optical phonon scattering is seen to increase much more quickly with temperature in GaAs

increases slightly with temperature for the case of GaN, from $6.7 \times 10^{13} \text{ s}^{-1}$ at 100 K to $8.6 \times 10^{13} \text{ s}^{-1}$ at 700 K, for high applied electric field strengths. Contrast this with the case of GaAs, where the rate increases from $4.0 \times 10^{12} \text{ s}^{-1}$ at 100 K to more than twice that amount at 700 K, $9.2 \times 10^{12} \text{ s}^{-1}$, at high applied electric field strengths. This large increase in the polar optical phonon scattering rate for the case of GaAs is one reason for the large drop in the electron drift velocity with increasing temperature for the case of GaAs.

A second reason for the variation in temperature dependence of the two materials is the occupancy of the upper valleys, shown in Fig. 33.8. In the case of GaN, the upper valleys begin to become occupied at roughly the same applied electric field strength, 100 kV/cm, independent of temperature. For the case of GaAs, however, the upper valleys are at a much lower energy than those in GaN. In particular, while the first upper conduction band valley minimum is 1.9 eV above the lowest point in the conduction band in GaN, the first upper conduc-

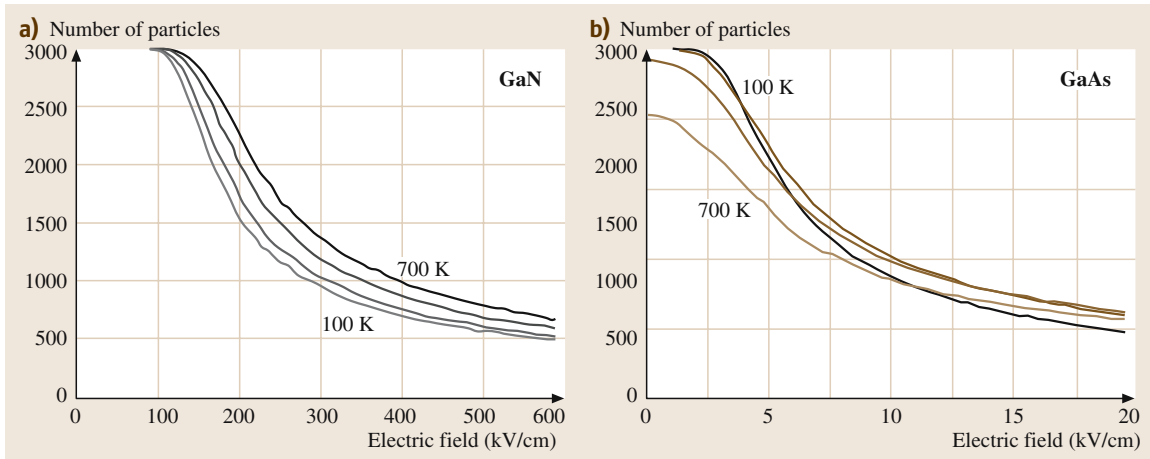


Fig. 33.8a,b A comparison of the number of particles in the lowest energy valley of the conduction band, the Γ valley, as a function of the applied electric field for various crystal temperatures, for the cases of (a) GaN and (b) GaAs. In GaAs, the electrons begin to occupy the upper valleys much more quickly, causing the electron drift velocity to drop as the crystal temperature is increased. Three thousand electrons were employed for these steady-state electron transport simulations

tion band valley is only 290 meV above the bottom of the conduction band in GaAs [33.51]. As the upper conduction band valleys are so close to the bottom of the conduction band for the case of GaAs, the thermal energy (at 700 K, $k_B T \simeq 60$ meV) is enough in order to allow for a small fraction of the electrons to transfer into the upper valleys even before an electric field is applied. When electrons occupy the upper valleys, intervalley scattering and the upper valleys' larger effective masses reduce the overall electron drift velocity. This

is another reason why the velocity–field characteristic associated with GaAs is more sensitive to variations in crystal temperature than that associated with GaN.

33.2.4 Influence of Doping on the Electron Drift Velocities Within GaN and GaAs

One parameter that can be readily controlled during the fabrication of semiconductor devices is the doping concentration. Understanding the effect of doping

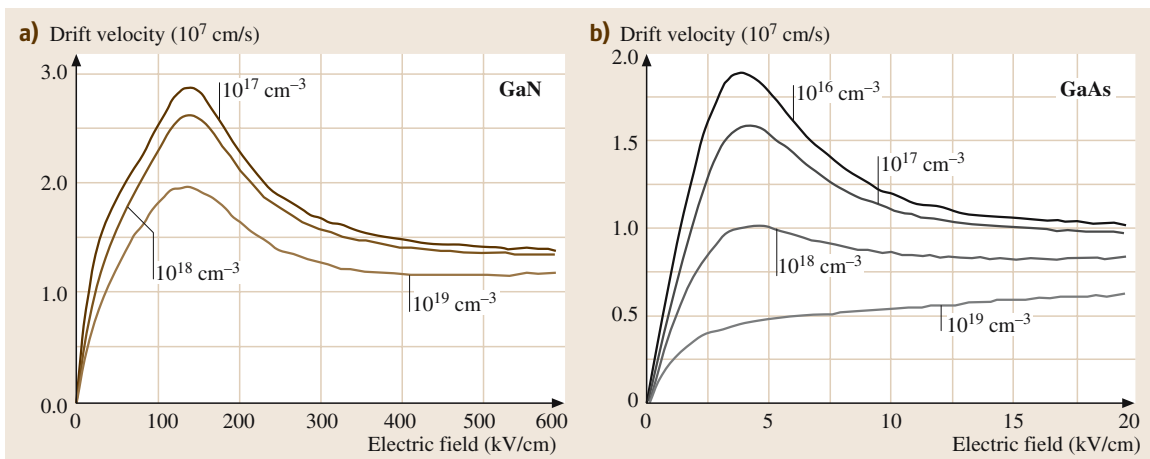


Fig. 33.9a,b A comparison of the dependence of the velocity–field characteristics associated with (a) GaN and (b) GaAs on the doping concentration. GaN maintains a higher electron drift velocity with increased doping levels than GaAs does

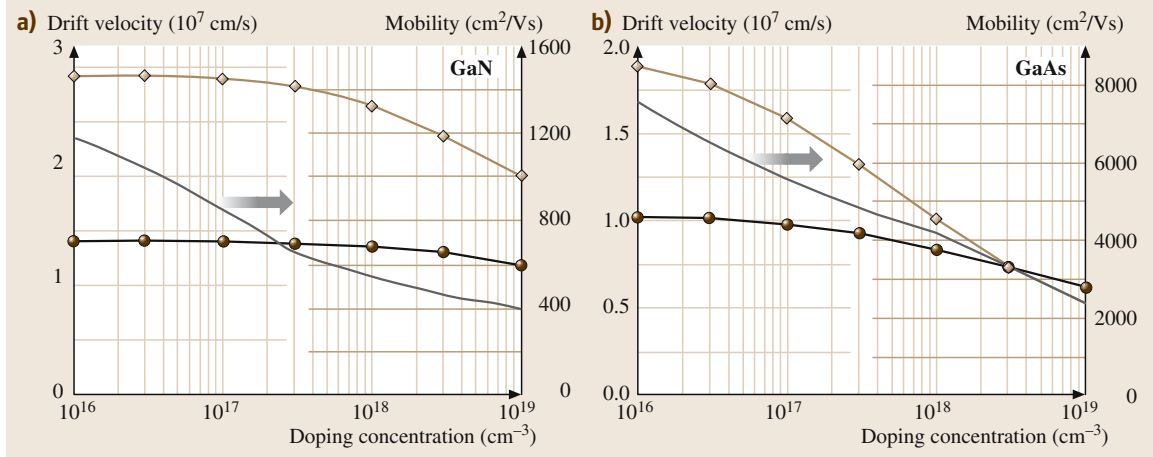


Fig. 33.10a,b A comparison of the low-field electron drift mobility (*solid lines*), the peak electron drift velocity (*diamonds*), and the saturation electron drift velocity (*solid points*) for **(a)** GaN and **(b)** GaAs as a function of the doping concentration. These parameters are more insensitive to increases in doping in GaN than in GaAs

on the resultant electron transport is also important. In Fig. 33.9, the velocity–field characteristic associated with GaN is presented for a number of different doping concentration levels. Once again, three important electron transport metrics are influenced by the doping concentration level: the low-field electron drift mobility, the peak electron drift velocity, and the saturation electron drift velocity; see Fig. 33.10. Our simulation results suggest that for doping concentrations of less than 10¹⁷ cm⁻³, there is very little effect on the velocity–field characteristic for the case of GaN. However, for doping concentrations above 10¹⁷ cm⁻³, the peak electron drift velocity diminishes considerably, from 2.9 × 10⁷ cm/s for the case of 10¹⁷ cm⁻³ doping to 2.0 × 10⁷ cm/s for the case of 10¹⁹ cm⁻³ doping. The saturation electron drift velocity within GaN is found to only decrease slightly in response to increases in the doping concentration. The effect of doping on the low-field electron drift mobility is also shown. It is seen that this mobility drops significantly in response to increases in the doping concentration level, from 1200 cm²/Vs at 10¹⁶ cm⁻³ doping to 400 cm²/Vs at 10¹⁹ cm⁻³ doping.

As we did for temperature, we compare the sensitivity of the velocity–field characteristic associated with GaN to doping with that associated with GaAs. Figure 33.9 shows this comparison. For the case of GaAs, it is seen that the electron drift velocities decrease much more with increased doping than those associated with GaN. In particular, for the case of GaAs, the peak electron drift velocity decreases from 1.8 × 10⁷ cm/s at 10¹⁶ cm⁻³ doping to 0.6 × 10⁷ cm/s at 10¹⁹ cm⁻³ dop-

ing. For GaAs, at the higher doping levels, the peak in the velocity–field characteristic disappears completely for sufficiently high doping concentrations. The saturation electron drift velocity decreases from 1.0 × 10⁷ cm/s at 10¹⁶ cm⁻³ doping to 0.6 × 10⁷ cm/s at 10¹⁹ cm⁻³ doping. The low-field electron drift mobility also diminishes dramatically with increased doping, dropping from 7800 cm²/Vs at 10¹⁶ cm⁻³ doping to 2200 cm²/Vs at 10¹⁹ cm⁻³ doping.

Once again, it is interesting to determine why the doping dependence in GaAs is so much more pronounced than it is in GaN. Again, we examine the polar optical phonon scattering rate and the occupancy of the upper valleys. Figure 33.11 shows the polar optical phonon scattering rates as a function of the applied electric field, for both GaN and GaAs. In this case, however, due to screening effects, the rate drops when the doping concentration is increased. The decrease, however, is much more pronounced for the case of GaAs than for GaN. It is believed that this drop in the polar optical phonon scattering rate allows for upper valley occupancy to occur more quickly in GaAs rather than in GaN (Fig. 33.12). For GaN, electrons begin to occupy the upper valleys at roughly the same applied electric field strength, independent of the doping level. However, for the case of GaAs, the upper valleys are occupied more quickly with greater doping. When the upper valleys are occupied, the electron drift velocity decreases due to intervalley scattering and the larger effective mass of the electrons within the upper valleys.

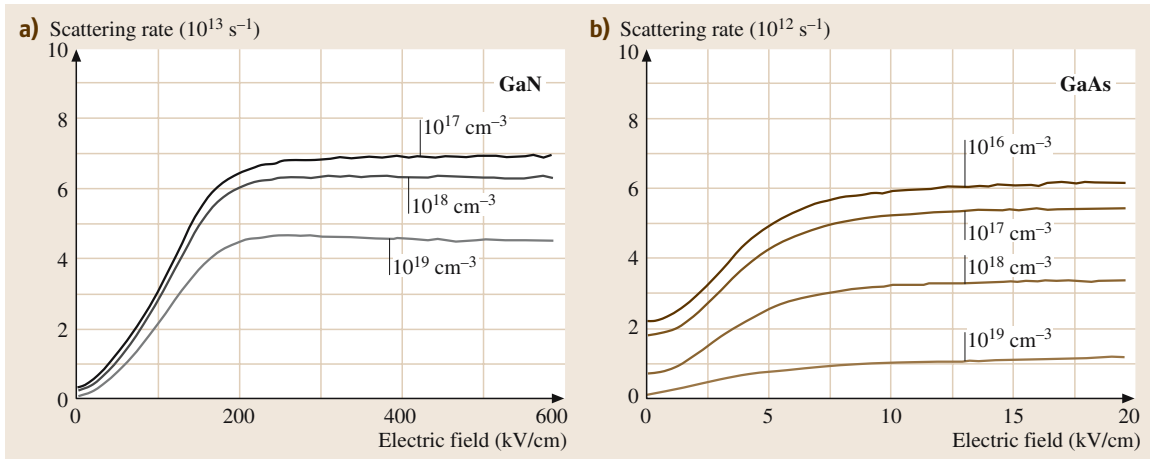


Fig. 33.11a,b A comparison of the polar optical phonon scattering rates as a function of the applied electric field, for both (a) GaN and (b) GaAs, for various doping concentrations

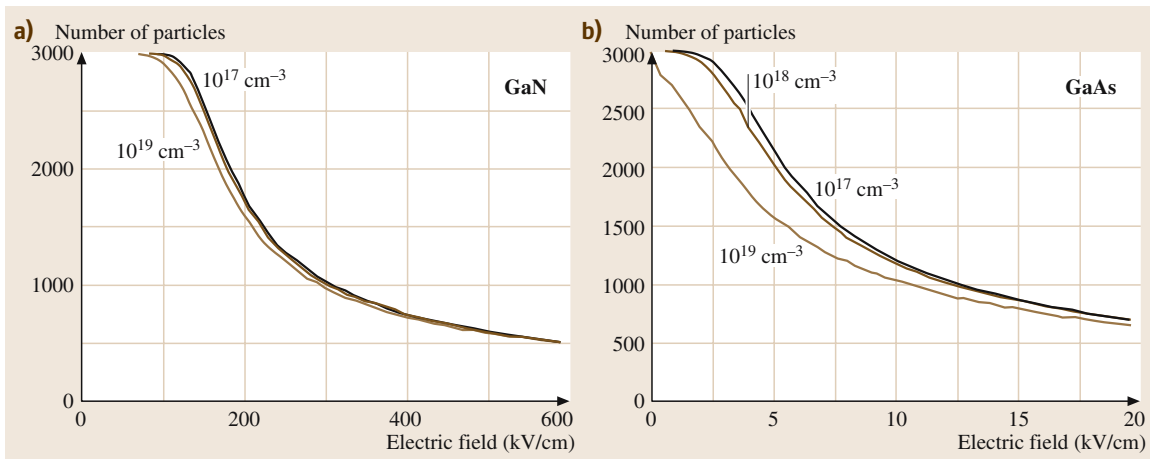


Fig. 33.12a,b A comparison of the number of particles in the lowest valley of the conduction band, the Γ valley, as a function of the applied electric field, for both (a) GaN and (b) GaAs, for various doping concentration levels. Three thousand electrons were employed for these steady-state electron transport simulations

33.2.5 Electron Transport in AlN

AlN has the largest effective mass of the III–V nitride semiconductors considered in this analysis. Accordingly, it is not surprising that this material exhibits the lowest electron drift velocity and the lowest low-field electron drift mobility. The sensitivity of the velocity–field characteristic associated with AlN to variations in the crystal temperature may be examined by considering Fig. 33.13. As with the case of GaN, the velocity–field characteristic associated with AlN is extremely robust to variations in the crystal tem-

perature. In particular, its peak electron drift velocity, which is $1.8 \times 10^7 \text{ cm/s}$ at 100 K, only decreases to $1.2 \times 10^7 \text{ cm/s}$ at 700 K. Similarly, its saturation electron drift velocity, which is $1.5 \times 10^7 \text{ cm/s}$ at 100 K, only decreases to $1.0 \times 10^7 \text{ cm/s}$ at 700 K. The low-field electron drift mobility associated with AlN also diminishes in response to increases in the crystal temperature, from $375 \text{ cm}^2/\text{Vs}$ at 100 K to $40 \text{ cm}^2/\text{Vs}$ at 700 K.

The sensitivity of the velocity–field characteristic associated with AlN to variations in the doping concentration may be examined by considering Fig. 33.14. It

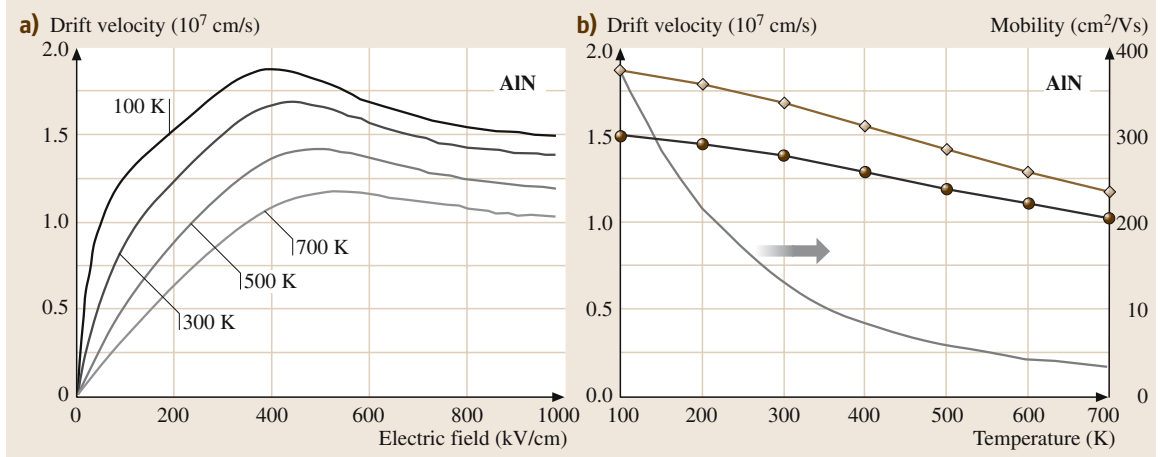


Fig. 33.13a,b The velocity–field characteristic associated with AlN (a) for various crystal temperatures. The trends in the low-field mobility (solid line), the peak electron drift velocity (diamonds), and the saturation electron drift velocity (solid points), are also shown. AlN exhibits its peak electron drift velocity at very high applied electric fields. AlN has the lowest peak electron drift velocity and the lowest low-field electron drift mobility of the III–V nitride semiconductors considered in this analysis (b)

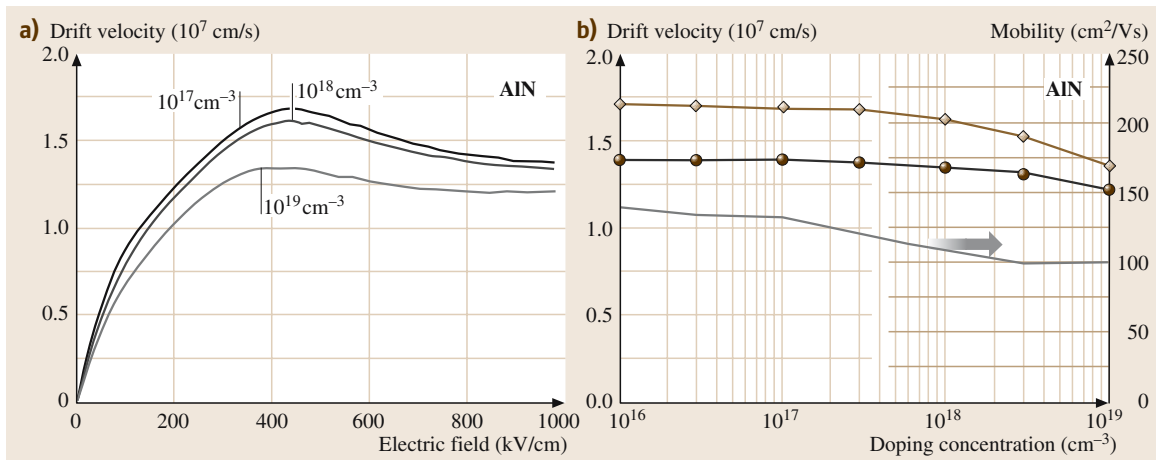


Fig. 33.14a,b The velocity–field characteristic associated with AlN for various doping concentrations (a). The trends in the low-field electron drift mobility (solid line), the peak electron drift velocity (diamonds), and the saturation electron drift velocity (solid points), are also shown (b)

is noted that the variations in the velocity–field characteristic associated with AlN in response to variations in the doping concentration are not as pronounced as those which occur in response to variations in the crystal temperature. Quantitatively, the peak electron drift velocity drops from 1.7×10^7 cm/s at 10^{17} cm^{-3} doping to 1.3×10^7 cm/s at 10^{19} cm^{-3} doping. Similarly, its saturation electron drift velocity drops from

1.4×10^7 cm/s at 10^{17} cm^{-3} doping to 1.2×10^7 cm/s at 10^{19} cm^{-3} doping. The influence of doping on the low-field electron drift mobility associated with AlN is also observed to be not as pronounced as for the case of crystal temperature. Figure 33.14b shows that the low-field electron drift mobility associated with AlN decreases from $140 \text{ cm}^2/\text{Vs}$ at 10^{16} cm^{-3} doping to $100 \text{ cm}^2/\text{Vs}$ at 10^{19} cm^{-3} doping.

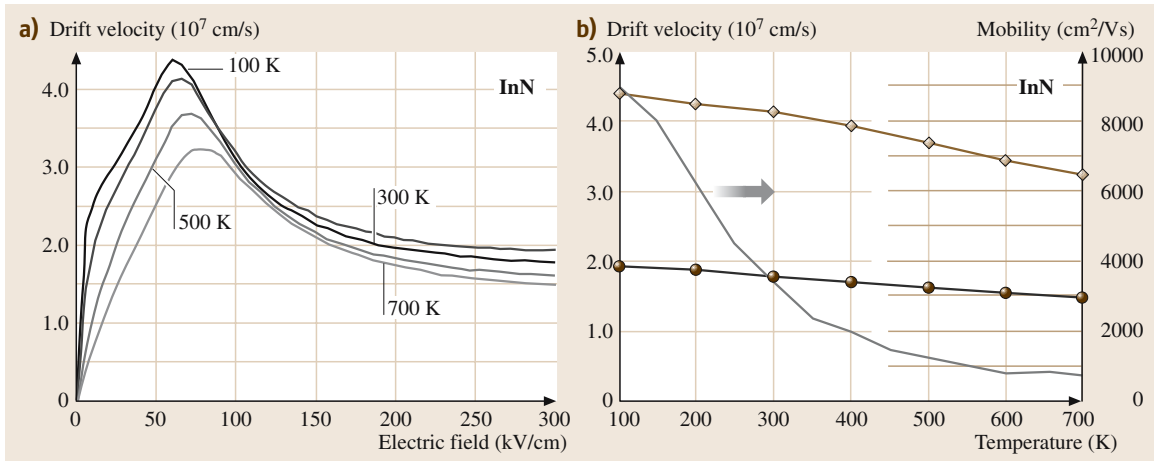


Fig. 33.15a,b The velocity–field characteristic associated with InN for various crystal temperatures (a). The trends in the low-field electron drift mobility (solid line), the peak electron drift velocity (diamonds), and the saturation electron drift velocity (solid points), are also shown. (b) InN has the highest peak electron drift velocity and the highest low-field electron drift mobility of the III–V nitride semiconductors considered in this analysis

33.2.6 Electron Transport in InN

InN has the smallest effective mass of the three III–V nitride semiconductors considered in this analysis. Accordingly, it is not surprising that it exhibits the highest electron drift velocity and the highest low-field electron drift mobility. The sensitivity of the velocity–field characteristic associated with InN to variations in the crystal temperature may be examined by consider-

ing Fig. 33.15. As with the cases of GaN and AlN, the velocity–field characteristic associated with InN is extremely robust to increases in the crystal temperature. In particular, its peak electron drift velocity, which is 4.4×10^7 cm/s at 100 K, only decreases to 3.2×10^7 cm/s at 700 K. Similarly, its saturation electron drift velocity, which is 2.0×10^7 cm/s at 100 K, only decreases to 1.5×10^7 cm/s at 700 K. The low-field electron drift mobility associated with InN also

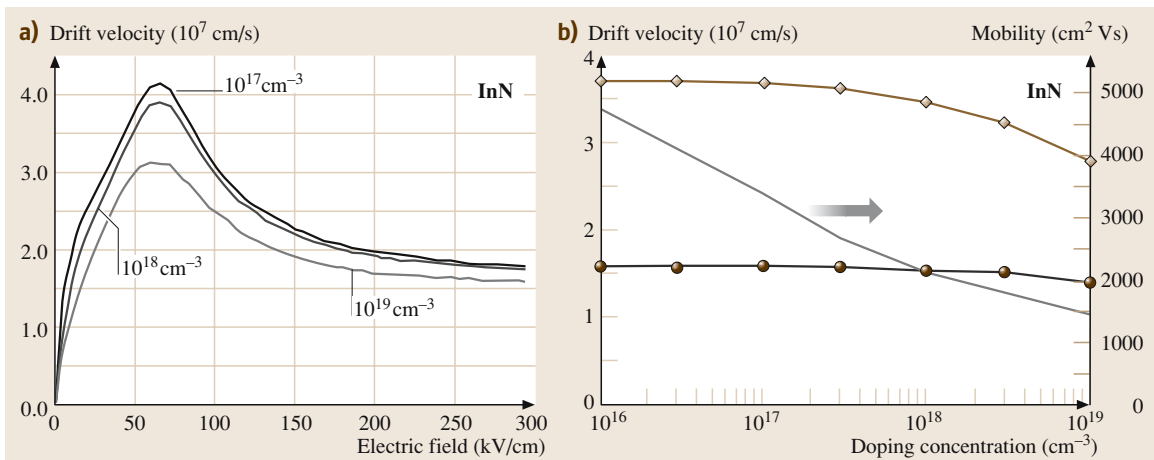


Fig. 33.16a,b The velocity–field characteristic associated with InN for various doping concentrations (a). The trends in the low-field electron drift mobility (solid line), the peak electron drift velocity (diamonds), and the saturation electron drift velocity (solid points), are also shown (b)

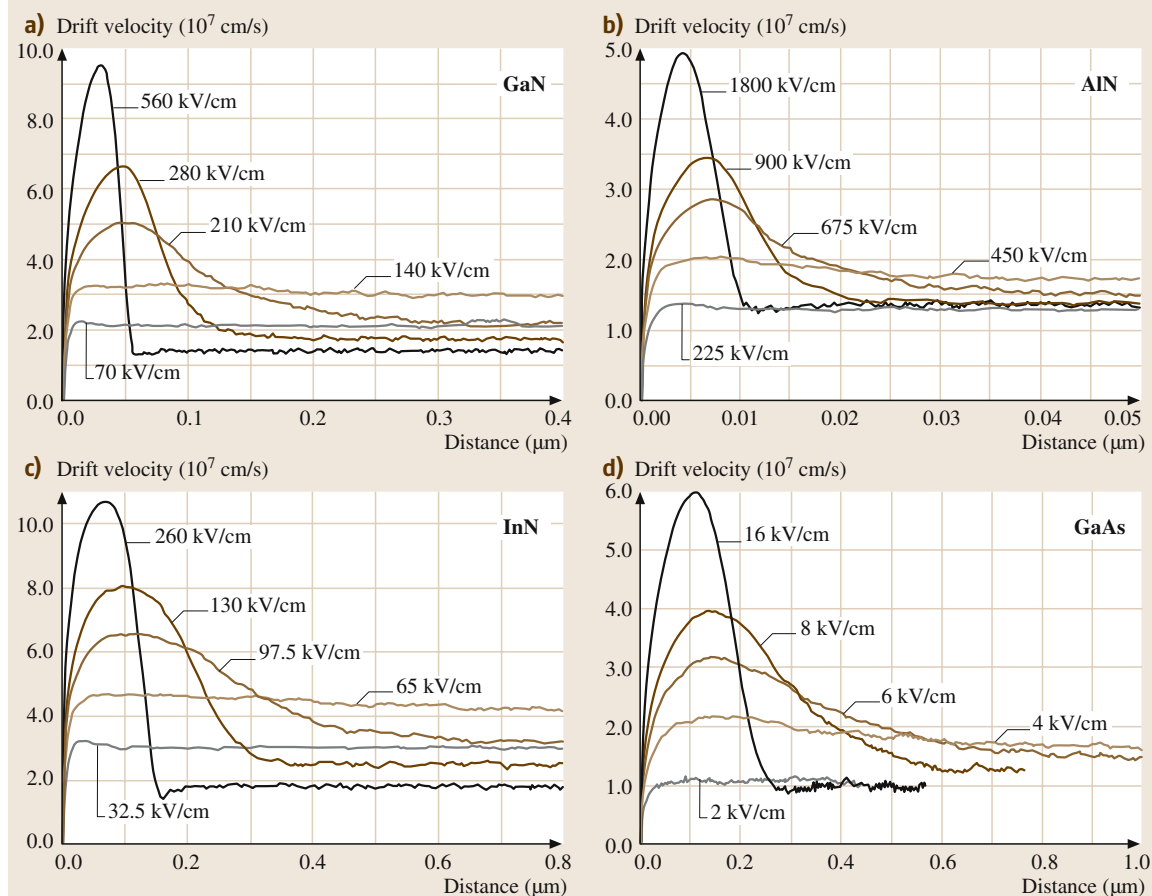


Fig. 33.17a–d The electron drift velocity as a function of the distance displaced since the application of the electric field for various applied electric field strengths, for the cases of (a) GaN, (b) AlN, (c) InN, and (d) GaAs. In all cases, we have assumed an initial zero field electron distribution, a crystal temperature of 300 K, and a doping concentration of 10^{17} cm^{-3} . After [33.29] with permission, copyright AIP

diminishes in response to increases in the crystal temperature, from about $9000 \text{ cm}^2/\text{Vs}$ at 100 K to below $1000 \text{ cm}^2/\text{Vs}$ at 700 K.

The sensitivity of the velocity–field characteristic associated with InN to variations in the doping concentration may be examined by considering Fig. 33.16. These results suggest a similar robustness to the doping concentration for the case of InN. In particular, it is noted that for doping concentrations below 10^{17} cm^{-3} , the velocity–field characteristic associated with InN exhibits very little dependence on the doping concentration. When the doping concentration is increased above 10^{17} cm^{-3} , however, the peak electron drift velocity diminishes. Quantitatively, the peak electron drift velocity decreases from $4.1 \times 10^7 \text{ cm/s}$ at 10^{17} cm^{-3} doping to

$3.2 \times 10^7 \text{ cm/s}$ at 10^{19} cm^{-3} doping. The saturation electron drift velocity only drops slightly, however, from $1.8 \times 10^7 \text{ cm/s}$ at 10^{17} cm^{-3} doping to $1.5 \times 10^7 \text{ cm/s}$ at 10^{19} cm^{-3} doping. The low-field electron drift mobility, however, drops significantly with doping, from $4700 \text{ cm}^2/\text{Vs}$ at 10^{16} cm^{-3} doping to $1500 \text{ cm}^2/\text{Vs}$ at 10^{19} cm^{-3} doping.

33.2.7 Transient Electron Transport

Steady-state electron transport is the dominant electron transport mechanism in devices with larger dimensions. For devices with smaller dimensions, however, transient electron transport must also be considered when evaluating device performance. *Ruch* [33.57] demonstrated,

for both silicon and GaAs, that the transient electron drift velocity may exceed the corresponding steady-state electron drift velocity by a considerable margin for appropriate selections of the applied electric field. *Shur and Eastman* [33.58] explored the device implications of transient electron transport, and demonstrated that substantial improvements in the device performance can be achieved as a consequence. *Heiblum et al.* [33.59] made the first direct experimental observation of transient electron transport within GaAs. Since then there have been a number of experimental investigations into the transient electron transport within the III–V compound semiconductors; see, for example, [33.60–62].

Thus far, very little research has been invested into the study of transient electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. *Foutz et al.* [33.21] examined transient electron transport within both the wurtzite and zinc blende phases of GaN. In particular, they examined how electrons, initially in thermal equilibrium, respond to the sudden application of a constant electric field. In devices with dimensions greater than $0.2\ \mu\text{m}$, they found that steady-state electron transport is expected to dominate device performance. For devices with smaller dimensions, however, upon the application of a sufficiently high electric field, they found that the transient electron drift velocity can considerably overshoot the corresponding steady-state electron drift velocity. This velocity overshoot was found to be comparable with that which occurs within GaAs.

Foutz et al. [33.29] performed a subsequent analysis in which the transient electron transport within all of the III–V nitride semiconductors under consideration in this analysis were compared with that which occurs within GaAs. In particular, following the approach of *Foutz et al.* [33.21], they examined how electrons, initially in thermal equilibrium, respond to the sudden application of a constant electric field. A key result of this study, presented in Fig. 33.17, plots the transient electron drift velocity as a function of the distance displaced since the electric field was initially applied for a number of applied electric field strengths and for each of the materials considered in this analysis.

Focusing initially on the case of GaN (Fig. 33.17a), we note that the electron drift velocity for the applied electric field strengths $70\ \text{kV/cm}$ and $140\ \text{kV/cm}$ reaches steady-state very quickly, with little or no velocity overshoot. In contrast, for applied electric field strengths above $140\ \text{kV/cm}$, significant velocity overshoot occurs. This result suggests that in GaN,

$140\ \text{kV/cm}$ is a critical field for the onset of velocity overshoot effects. As mentioned in Sect. 33.2.2, $140\ \text{kV/cm}$ also corresponds to the peak in the velocity-field characteristic associated with GaN; recall Fig. 33.4. Steady-state Monte Carlo simulations suggest that this is the point at which significant upper valley occupation begins to occur; recall Fig. 33.3. This suggests that velocity overshoot is related to the transfer of electrons to the upper valleys. Similar results are found for the other III–V nitride semiconductors, AlN and InN, and GaAs; see Figs. 33.17b–d.

We now compare the transient electron transport characteristics for the materials. From Fig. 33.17, it is clear that certain materials exhibit higher peak overshoot velocities and longer overshoot relaxation times. It is not possible to fairly compare these different semiconductors by applying the same applied electric field strength to each of the materials, as the transient effects occur over such a disparate range of applied electric field strengths in each material. In order to facilitate such a comparison, we choose a field strength equal to twice the critical applied electric field strength for each material. Figure 33.18 shows a comparison of the velocity overshoot effects amongst the four materials considered in this analysis, i.e., GaN, AlN, InN, and

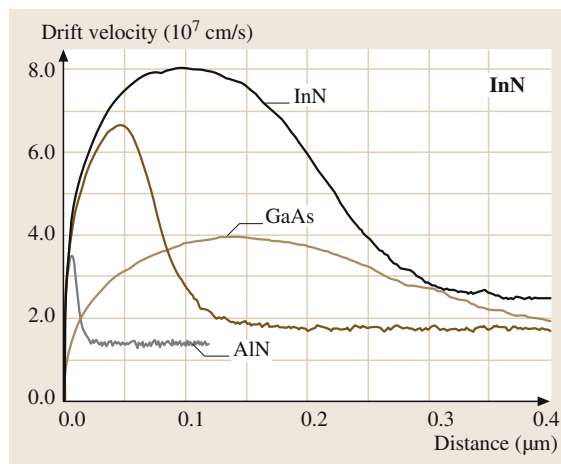


Fig. 33.18 A comparison of the velocity overshoot amongst the III–V nitride semiconductors and GaAs. The applied electric field strength chosen corresponds to twice the critical applied electric field strength at which the peak in the steady-state velocity–field characteristic occurs (Fig. 33.4), i.e., $280\ \text{kV/cm}$ for the case of GaN, $900\ \text{kV/cm}$ for the case of AlN, $130\ \text{kV/cm}$ for the case of InN, and $8\ \text{kV/cm}$ for the case of GaAs. After [33.29] with permission, copyright AIP

GaAs. It is clear that among the three III–V nitride semiconductors considered, InN exhibits superior transient electron transport characteristics. In particular, InN has the largest overshoot velocity and the distance over which this overshoot occurs, $0.3\text{ }\mu\text{m}$, is longer than in either GaN and AlN. GaAs exhibits a longer overshoot relaxation distance, approximately $0.7\text{ }\mu\text{m}$, but the electron drift velocity exhibited by InN is greater than that of GaAs for all distances.

33.2.8 Electron Transport: Conclusions

In this section, steady-state and transient electron transport results, corresponding to the III–V nitride semiconductors, GaN, AlN, and InN, were presented, these results being obtained from our Monte Carlo simulations of the electron transport within these materials.

Steady-state electron transport was the dominant theme of our analysis. In order to aid in the understanding of these electron transport characteristics, a comparison was made between GaN and GaAs. Our simulations showed that GaN is more robust to variations in crystal temperature and doping concentration than GaAs, and an analysis of our Monte Carlo simulation results showed that polar optical phonon scattering plays the dominant role in accounting for these differences in behavior. This analysis was also performed for the other III–V nitride semiconductors considered in this analysis – AlN and InN – and similar results were obtained. Finally, we presented some key transient electron transport results. These results indicated that the transient electron transport that occurs within InN is the most pronounced of all of the materials under consideration in this review (GaN, AlN, InN, and GaAs).

33.3 Electron Transport Within III–V Nitride Semiconductors: A Review

Pioneering investigations into the material properties of the III–V nitride semiconductors, GaN, AlN, and InN, were performed during the earlier half of the twentieth century [33.63–65]. The III–V nitride semiconductor materials available at the time, small crystals and powders, were of poor quality, and completely unsuitable for device applications. Thus, it was not until the late 1960s, when *Maruska* and *Tietjen* [33.66] employed chemical vapor deposition to fabricate GaN, that interest in the III–V nitride semiconductors experienced a renaissance. Since that time, interest in the III–V nitride semiconductors has been growing, the material properties of these semiconductors improving considerably over the years. As a result of this research effort, there are currently a number of commercial devices available that employ the III–V nitride semiconductors. More III–V nitride semiconductor-based device applications are currently under development, and these should become available in the near future.

In this section, we present a brief overview of the III–V nitride semiconductor electron transport field. We start with a survey describing the evolution of the field. In particular, the sequence of critical developments that have occurred that contribute to our current understanding of the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, is chronicled. Then, some of the current literature is presented, with particular emphasis being placed on the most recent developments in the field and how such

developments are modifying our understanding of the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN. Finally, frontiers for further research and investigation are presented.

This section is organized in the following manner. In Sect. 33.3.1, we present a brief survey describing the evolution of the field. Then, in Sect. 33.3.2, the current literature is discussed. Finally, frontiers for further research and investigation are presented in Sect. 33.3.3.

33.3.1 Evolution of the Field

The favorable electron transport characteristics of the III–V nitride semiconductors, GaN, AlN, and InN, have long been recognized. As early as the 1970s, *Littlejohn et al.* [33.13] pointed out that the large polar optical phonon energy characteristic of GaN, in conjunction with its large intervalley energy separation, suggests a high saturation electron drift velocity for this material. As the high-frequency electron device performance is, to a large degree, determined by this saturation electron drift velocity [33.14], the recognition of this fact ignited enhanced interest in this material and its III–V nitride semiconductor compatriots, AlN and InN. This enhanced interest, and the developments which have transpired as a result of it, are responsible for the III–V nitride semiconductor industry of today.

In 1975, *Littlejohn et al.* [33.13] were the first to report results obtained from semi-classical Monte Carlo

simulations of the steady-state electron transport within bulk wurtzite GaN. A one-valley model for the conduction band was adopted in their analysis. Steady-state electron transport, for both parabolic and nonparabolic band structures, was considered in their analysis, nonparabolicity being treated through the application of the Kane model [33.43]. The primary focus of their investigation was the determination of the velocity-field characteristic associated with GaN. All donors were assumed to be ionized, and the free electron concentration was taken to be equal to the dopant concentration. The scattering mechanisms considered were (1) ionized impurity, (2) polar optical phonon, (3) piezoelectric, and (4) acoustic deformation potential. For the case of the parabolic band, in the absence of ionized impurities, they found that the electron drift velocity monotonically increases with the applied electric field strength, saturating at a value of about 2.5×10^7 cm/s for the case of high applied electric fields. In contrast, for the case of the nonparabolic band, and in the absence of ionized impurities, a region of negative differential mobility was found, the electron drift velocity achieving a maximum of about 2×10^7 cm/s at an applied electric field strength of about 100 kV/cm, with further increases in the applied electric field strength resulting in a slight decrease in the corresponding electron drift velocity. The role of ionized impurity scattering was also investigated by *Littlejohn et al.* [33.13].

In 1993, *Gelmont et al.* [33.16] reported on ensemble semi-classical two-valley Monte Carlo simulations of the electron transport within bulk wurtzite GaN, this analysis improving upon the analysis of *Littlejohn et al.* [33.13] by incorporating intervalley scattering into the simulations. They found that the negative differential mobility found in bulk wurtzite GaN is much more pronounced than that found by *Littlejohn et al.* [33.13], and that intervalley transitions are responsible for this. For a doping concentration of 10^{17} cm⁻³, *Gelmont et al.* [33.16] demonstrated that the electron drift velocity achieves a peak value of about 2.8×10^7 cm/s at an applied electric field of about 140 kV/cm. The impact of intervalley transitions on the electron distribution function was also determined and shown to be significant. The impact of doping and compensation on the velocity-field characteristic associated with bulk wurtzite GaN was also examined.

Since these pioneering investigations, ensemble Monte Carlo simulations of the electron transport within GaN have been performed numerous times. In particular, in 1995 *Mansour et al.* [33.18] reported the use of such an approach in order to determine how the crystal

temperature influences the velocity-field characteristic associated with bulk wurtzite GaN. Later that year, *Kolník et al.* [33.19] reported on employing full-band Monte Carlo simulations of the electron transport within bulk wurtzite GaN and bulk zinc blende GaN, finding that bulk zinc blende GaN exhibits a much higher low-field electron drift mobility than bulk wurtzite GaN. The peak electron drift velocity corresponding to bulk zinc blende GaN was found to be only marginally greater than that exhibited by bulk wurtzite GaN. In 1997, *Bhaskar and Shur* [33.22] reported on employing ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within bulk and confined wurtzite GaN. Their simulations demonstrated that the two-dimensional electron gas within a confined wurtzite GaN structure will exhibit a higher low-field electron drift mobility than bulk wurtzite GaN, by almost an order of magnitude, this being in agreement with experiment. In 1998, *Albrecht et al.* [33.27] reported on employing ensemble semi-classical five-valley Monte Carlo simulations of the electron transport within bulk wurtzite GaN, with the aim of determining elementary analytical expressions for a number of electron transport metrics corresponding to bulk wurtzite GaN, for the purposes of device modeling.

Electron transport within the other III–V nitride semiconductors, AlN and InN, has also been studied using ensemble semi-classical Monte Carlo simulations of the electron transport. In particular, by employing ensemble semi-classical three-valley Monte Carlo simulations, the velocity-field characteristic associated with bulk wurtzite AlN was studied and reported by *O'Leary et al.* [33.24] in 1998. They found that AlN exhibits the lowest peak and saturation electron drift velocities of the III–V nitride semiconductors considered in this analysis. Similar simulations of the electron transport within bulk wurtzite AlN were also reported by *Albrecht et al.* [33.25] in 1998. The results of *O'Leary et al.* [33.24] and *Albrecht et al.* [33.25] were found to be quite similar. The first known simulation of the electron transport within bulk wurtzite InN was the semi-classical three-valley Monte Carlo simulation of *O'Leary et al.* [33.23], reported in 1998. InN was demonstrated to have the highest peak and saturation electron drift velocities of the III–V nitride semiconductors. The subsequent ensemble full-band Monte Carlo simulations of *Bellotti et al.* [33.28], reported in 1999, produced results similar to those of *O'Leary et al.* [33.23].

The first known study of transient electron transport within the III–V nitride semiconductors was that performed by *Foutz et al.* [33.21], reported in 1997. In

this study, ensemble semi-classical three-valley Monte Carlo simulations were employed in order to determine how the electrons within wurtzite and zinc blende GaN, initially in thermal equilibrium, respond to the sudden application of a constant electric field. The velocity overshoot that occurs within these materials was examined. It was found that the electron drift velocities that occur within the zinc blende phase of GaN are slightly greater than those exhibited by the wurtzite phase owing to the slightly higher steady-state electron drift velocity exhibited by the zinc blende phase of GaN. A comparison with the transient electron transport that occurs within GaAs was made. Using the results from this analysis, a determination of the minimum transit time as a function of the distance displaced since the application of the applied electric field was performed for all three materials considered in this study: wurtzite GaN, zinc blende GaN, and GaAs. For distances in excess of 0.1 μm , both phases of GaN were shown to exhibit superior performance (reduced transit time) when contrasted with that associated with GaAs.

A more general analysis, in which transient electron transport within GaN, AlN, and InN was studied, was performed by Foutz et al. [33.29], and reported in 1999. As with their previous study, Foutz et al. [33.29] determined how electrons, initially in thermal equilibrium, respond to the sudden application of a constant electric field. For GaN, AlN, InN, and GaAs, it was found that the electron drift velocity overshoot only occurs when the applied electric field exceeds a certain critical applied electric field strength unique to each material. The critical applied electric field strengths, 140 kV/cm for the case of wurtzite GaN, 450 kV/cm for the case of AlN, 65 kV/cm for the case of InN, and 4 kV/cm for the case of GaAs, were shown to correspond to the peak electron drift velocity in the velocity-field characteristic associated with each of these materials; recall Fig. 33.4. It was found that InN exhibits the highest peak overshoot velocity, and that this overshoot lasts over prolonged distances compared with AlN, InN, and GaAs. A comparison with the results of experiment was performed.

In addition to Monte Carlo simulations of the electron transport within these materials, a number of other types of electron transport studies have been performed. In 1975, for example, Ferry [33.14] reported on the determination of the velocity-field characteristic associated with wurtzite GaN using a displaced Maxwellian distribution function approach. For high applied electric fields, Ferry [33.14] found that the electron drift velocity associated with GaN monotonically increases

with the applied electric field strength (it does not saturate), reaching a value of about 2.5×10^7 cm/s at an applied electric field strength of 300 kV/cm. The device implications of this result were further explored by Das and Ferry [33.15]. In 1994, Chin et al. [33.17] reported on a detailed study of the dependence of the low-field electron drift mobilities associated with the III–V nitride semiconductors, GaN, AlN, and InN, on crystal temperature and doping concentration. An analytical expression for the low-field electron drift mobility, μ , determined using a variational principle, was used for the purposes of this analysis. The results obtained were contrasted with those from experiment. A subsequent mobility study was reported in 1997 by Look et al. [33.38]. Then, in 1998, Weimann et al. [33.26] reported on a model for determining how the scattering of electrons by the threading dislocations within bulk wurtzite GaN influence the low-field electron drift mobility. They demonstrated why the experimentally measured low-field electron drift mobility associated with this material is much lower than that predicted from Monte Carlo analyses: threading dislocations were not taken into account in the Monte Carlo simulations of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN.

While the negative differential mobility exhibited by the velocity-field characteristics associated with the III–V nitride semiconductors, GaN, AlN, and InN, is widely attributed to intervalley transitions, and while direct experimental evidence confirming this has been presented [33.67], Krishnamurthy et al. [33.34] suggest that the inflection points in the bands located in the vicinity of the Γ valley are primarily responsible for the negative differential mobility exhibited by wurtzite GaN instead. The relative importance of these two mechanisms (intervalley transitions and inflection point considerations) were evaluated by Krishnamurthy et al. [33.34], for both bulk wurtzite GaN and an AlGaIn alloy.

33.3.2 Recent Developments

There have been a number of interesting recent developments in the study of the electron transport within the III–V nitride semiconductors which have influenced the direction of thought in this field. On the experimental front, in 2000 Wraback et al. [33.33] reported on the use of a femtosecond optically detected time-of-flight experimental technique in order to experimentally determine the velocity-field characteristic associated with bulk wurtzite GaN. They found that the peak electron

drift velocity, 1.9×10^7 cm/s, is achieved at an applied electric field strength of 225 kV/cm. No discernible negative differential mobility was observed. *Wraback et al.* [33.33] suggested that the large defect density characteristic of the GaN samples they employed, which were not taken into account in Monte Carlo simulations of the electron transport within this material, accounts for the difference between this experimental result and that obtained using simulation. They also suggested that decreasing the intervalley energy separation from about 2 eV to 340 meV, as suggested by the experimental results of *Brazel et al.* [33.56], may also account for these observations.

The determination of the electron drift velocity from experimental measurements of the unity gain cut-off frequency, f_t , has been pursued by a number of researchers. The key challenge in these analyses is the de-embedding of the parasitics from the experimental measurements so that the true intrinsic saturation electron drift velocity may be obtained. *Eastman et al.* [33.68] present experimental evidence which suggests that the saturation electron drift velocity within bulk wurtzite GaN is about 1.2×10^7 – 1.3×10^7 cm/s. A more recent report, by *Oxley and Uren* [33.69], suggests a value of 1.1×10^7 cm/s. The role of self-heating was also probed by *Oxley and Uren* [33.69] and shown to be relatively insignificant. A completely satisfactory explanation for the discrepancy between these results and those from the Monte Carlo simulations has yet to be provided.

Wraback et al. [33.70] performed a subsequent study on the transient electron transport within wurtzite GaN. In particular, using their femtosecond optically detected time-of-flight experimental technique in order to experimentally determine the velocity overshoot that occurs within bulk wurtzite GaN, they observed substantial velocity overshoot within this material. In particular, a peak transient electron drift velocity of 7.25×10^7 cm/s was observed within the first 200 fs after photoexcitation for an applied electric field strength of 320 kV/cm. These experimental results were shown to be consistent with the theoretical predictions of *Foutz et al.* [33.29].

On the theoretical front, there have been a number of recent developments. In 2001, *O'Leary et al.* [33.30] presented an elementary, one-dimensional analytical model for the electron transport within the III–V compound semiconductors, and applied it to the cases of wurtzite GaN and GaAs. The predictions of this analytical model were compared with those of Monte Carlo simulations and were found to be in satisfactory agreement. Hot-electron energy relaxation times within the III–V nitride semiconductors were recently studied by *Matulionis*

et al. [33.71] and reported in 2002. *Bulutay et al.* [33.72] studied the electron momentum and energy relaxation times within the III–V nitride semiconductors and reported the results of this study in 2003. It is particularly interesting to note that their arguments add considerable credence to the earlier inflection point argument of *Krishnamurthy et al.* [33.34]. In 2004, *Brazis and Raguotis* [33.73] reported on the results of a Monte Carlo study involving additional phonon modes and a smaller intervalley energy separation for bulk wurtzite GaN. Their results were found to be much closer to the experimental results of *Wraback et al.* [33.33] than those found previously.

The influence of hot-phonons on the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, has been the focus of considerable recent investigation. In particular, in 2004 *Silva and Nascimento* [33.74], *Gökden* [33.75], and *Ridley et al.* [33.76], to name just three, presented results related to this research focus. These results suggest that hot-phonon effects play a significant role in influencing the nature of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN. In particular, *Ridley et al.* [33.76] point out that the saturation electron drift velocity and the applied electric field strength at which the peak in the velocity–field characteristic occurs are both influenced by hot-phonon effects. The role that hot-phonons play in influencing device performance was studied by *Matulionis and Liberis* [33.77]. Research into the role that hot-phonons play in influencing the electron transport mechanisms within the III–V nitride semiconductors, GaN, AlN, and InN, seems likely to continue into the foreseeable future.

33.3.3 Future Perspectives

It is clear that our understanding of the electron transport within the III–V nitride semiconductors, GaN, AlN, and InN, is, at present at least, in a state of flux. A complete understanding of the electron transport mechanisms within these materials has yet to be achieved, and is the subject of intense current research. Most troubling is the discrepancy between the results of experiment and those of simulation. There are two principal sources of uncertainty in our analysis of the electron transport mechanisms within these materials: (1) uncertainty in the material properties, and (2) uncertainty in the underlying physics. We discuss each of these subsequently.

Uncertainty in the material parameters associated with the III–V nitride semiconductors, GaN, AlN, and

InN, remains a key source of ambiguity in the analysis of the electron transport with these materials [33.32]. Even for bulk wurtzite GaN, the most studied of the III–V nitride semiconductors considered in this analysis, uncertainty in the band structure remains an issue [33.56]. The energy gap associated with InN and the effective mass associated with this material continue to fuel debate; see, for example, *Davydov et al.* [33.53], *Wu et al.* [33.54], and *Matsuoka et al.* [33.55]. Variations in the experimentally determined energy gap associated with InN, observed from sample to sample, further confound matters. Most recently, *Shubina et al.* [33.78] suggested that nonstoichiometry within InN may be responsible for these variations in the energy gap. Further research will have to be performed in order to confirm this. Given this uncertainty in the band structures associated with the III–V nitride semiconductors, it is

clear that new simulations of the electron transport will have to be performed once researchers have settled on the appropriate band structures. We thus view our present results as a baseline, the sensitivity analysis of *O’Leary et al.* [33.32] providing some insights into how variations in the band structures will impact upon the results.

Uncertainty in the underlying physics is considerable. The source of the negative differential mobility remains a matter to be resolved. The presence of hot-phonons within these materials, and how such phonons impact upon the electron transport mechanisms within these materials, remains another point of contention. It is clear that a deeper understanding of these electron transport mechanisms will have to be achieved in order for the next generation of III–V nitride semiconductor-based devices to be properly designed.

33.4 Conclusions

In this paper, we reviewed analyses of the electron transport within the III–V nitride semiconductors GaN, AlN, and InN. In particular, we have discussed the evolution of the field, surveyed the current literature, and presented frontiers for further investigation and analysis. In order to narrow the scope of this review, we focused on the electron transport within bulk wurtzite GaN, AlN, and InN for the purposes of this paper. Most of our discussion focused upon results obtained from our ensemble semi-classical three-valley Monte Carlo simulations of the electron transport within these materials, our results conforming with state-of-the-art III–V nitride semiconductor orthodoxy.

We began our review with the Boltzmann transport equation, since this equation underlies most analy-

ses of the electron transport within semiconductors. A brief description of our ensemble semi-classical three-valley Monte Carlo simulation approach to solving the Boltzmann transport equation was then provided. The material parameters, corresponding to bulk wurtzite GaN, AlN, and InN, were then presented. We then used these material parameter selections, and our ensemble semi-classical three-valley Monte Carlo simulation approach, to determine the nature of the steady-state and transient electron transport within the III–V nitride semiconductors. Finally, we presented some recent developments on the electron transport within these materials, and pointed to fertile frontiers for further research and investigation.

References

- 33.1 S. Strite, H. Morkoç: *J. Vac. Sci. Technol. B* **10**, 1237 (1992)
- 33.2 H. Morkoç, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, M. Burns: *J. Appl. Phys.* **76**, 1363 (1994)
- 33.3 S. N. Mohammad, H. Morkoç: *Prog. Quantum Electron.* **20**, 361 (1996)
- 33.4 S. J. Pearton, J. C. Zolper, R. J. Shul, F. Ren: *J. Appl. Phys.* **86**, 1 (1999)
- 33.5 M. A. Khan, J. W. Yang, W. Knap, E. Frayssinet, X. Hu, G. Simin, P. Prystawko, M. Leszczynski, I. Grzegory, S. Porowski, R. Gaska, M. S. Shur, B. Beaumont, M. Teisseire, G. Neu: *Appl. Phys. Lett.* **76**, 3807 (2000)
- 33.6 X. Hu, J. Deng, N. Pala, R. Gaska, M. S. Shur, C. Q. Chen, J. Yang, G. Simin, M. A. Khan, J. C. Rojo, L. J. Schowalter: *Appl. Phys. Lett.* **82**, 1299 (2003)
- 33.7 W. Lu, V. Kumar, E. L. Piner, I. Adesida: *IEEE Trans. Electron Dev.* **50**, 1069 (2003)
- 33.8 A. Jiménez, Z. Bougrioua, J. M. Tirado, A. F. Braña, E. Calleja, E. Muñoz, I. Moerman: *Appl. Phys. Lett.* **82**, 4827 (2003)

- 33.9 A.A. Burk Jr., M.J. O'Loughlin, R.R. Siergiej, A.K. Agarwal, S. Sriram, R.C. Clarke, M.F. MacMillan, V. Balakrishna, C.D. Brandt: *Solid-State Electron.* **43**, 1459 (1999)
- 33.10 M. Umeno, T. Egawa, H. Ishikawa: *Mater. Sci. Semicond. Process.* **4**, 459 (2001)
- 33.11 A. Krost, A. Dadgar: *Phys. Status Solidi A* **194**, 361 (2002)
- 33.12 C. L. Tseng, M. J. Youh, G. P. Moore, M. A. Hopkins, R. Stevens, W. N. Wang: *Appl. Phys. Lett.* **83**, 3677 (2003)
- 33.13 M. A. Littlejohn, J. R. Hauser, T. H. Glisson: *Appl. Phys. Lett.* **26**, 625 (1975)
- 33.14 D. K. Ferry: *Phys. Rev. B* **12**, 2361 (1975)
- 33.15 P. Das, D. K. Ferry: *Solid-State Electron.* **19**, 851 (1976)
- 33.16 B. Gelmont, K. Kim, M. Shur: *J. Appl. Phys.* **74**, 1818 (1993)
- 33.17 V. W. L. Chin, T. L. Tansley, T. Osotchan: *J. Appl. Phys.* **75**, 7365 (1994)
- 33.18 N. S. Mansour, K. W. Kim, M. A. Littlejohn: *J. Appl. Phys.* **77**, 2834 (1995)
- 33.19 J. Kolník, İ. H. Oğuzman, K. F. Brennan, R. Wang, P. P. Ruden, Y. Wang: *J. Appl. Phys.* **78**, 1033 (1995)
- 33.20 M. Shur, B. Gelmont, M. A. Khan: *J. Electron. Mater.* **25**, 777 (1996)
- 33.21 B. E. Foutz, L. F. Eastman, U. V. Bhapkar, M. S. Shur: *Appl. Phys. Lett.* **70**, 2849 (1997)
- 33.22 U. V. Bhapkar, M. S. Shur: *J. Appl. Phys.* **82**, 1649 (1997)
- 33.23 S. K. O'Leary, B. E. Foutz, M. S. Shur, U. V. Bhapkar, L. F. Eastman: *J. Appl. Phys.* **83**, 826 (1998)
- 33.24 S. K. O'Leary, B. E. Foutz, M. S. Shur, U. V. Bhapkar, L. F. Eastman: *Solid State Commun.* **105**, 621 (1998)
- 33.25 J. D. Albrecht, R. P. Wang, P. P. Ruden, M. Farahmand, K. F. Brennan: *J. Appl. Phys.* **83**, 1446 (1998)
- 33.26 N. G. Weimann, L. F. Eastman, D. Doppalapudi, H. M. Ng, T. D. Moustakas: *J. Appl. Phys.* **83**, 3656 (1998)
- 33.27 J. D. Albrecht, R. P. Wang, P. P. Ruden, M. Farahmand, K. F. Brennan: *J. Appl. Phys.* **83**, 4777 (1998)
- 33.28 E. Bellotti, B. K. Doshi, K. F. Brennan, J. D. Albrecht, P. P. Ruden: *J. Appl. Phys.* **85**, 916 (1999)
- 33.29 B. E. Foutz, S. K. O'Leary, M. S. Shur, L. F. Eastman: *J. Appl. Phys.* **85**, 7727 (1999)
- 33.30 S. K. O'Leary, B. E. Foutz, M. S. Shur, L. F. Eastman: *Solid State Commun.* **118**, 79 (2001)
- 33.31 T. F. de Vasconcelos, F. F. Maia Jr., E. W. S. Caetano, V. N. Freire, J. A. P. da Costa, E. F. da Silva Jr.: *J. Cryst. Growth* **246**, 320 (2002)
- 33.32 S. K. O'Leary, B. E. Foutz, M. S. Shur, L. F. Eastman: *J. Electron. Mater.* **32**, 327 (2003)
- 33.33 M. Wraback, H. Shen, J. C. Carrano, T. Li, J. C. Campbell, M. J. Schurman, I. T. Ferguson: *Appl. Phys. Lett.* **76**, 1155 (2000)
- 33.34 S. Krishnamurthy, M. van Schilfgaarde, A. Sher, A.-B. Chen: *Appl. Phys. Lett.* **71**, 1999 (1997)
- 33.35 D. K. Ferry, C. Jacoboni (Eds.): *Quantum Transport in Semiconductors* (Plenum, New York 1992)
- 33.36 N. W. Ashcroft, N. D. Mermin: *Solid State Physics* (Saunders College, Philadelphia 1976)
- 33.37 P. A. Sandborn, A. Rao, P. A. Blakey: *IEEE Trans. Electron Dev.* **36**, 1244 (1989)
- 33.38 D. C. Look, J. R. Sizelove, S. Keller, Y. F. Wu, U. K. Mishra, S. P. DenBaars: *Solid State Commun.* **102**, 297 (1997)
- 33.39 N. A. Zakhleniuk, C. R. Bennett, B. K. Ridley, M. Babiker: *Appl. Phys. Lett.* **73**, 2485 (1998)
- 33.40 B. R. Nag: *Electron Transport in Compound Semiconductors* (Springer, Berlin, Heidelberg 1980)
- 33.41 M. S. Krishnan, N. Goldsman, A. Christou: *J. Appl. Phys.* **83**, 5896 (1998)
- 33.42 R. Oberhuber, G. Zandler, P. Vogl: *Appl. Phys. Lett.* **73**, 818 (1998)
- 33.43 W. Fawcett, A. D. Boardman, S. Swain: *J. Phys. Chem. Solids* **31**, 1963 (1970)
- 33.44 A. Bykhovski, B. Gelmont, M. Shur, A. Khan: *J. Appl. Phys.* **77**, 1616 (1995)
- 33.45 A. D. Bykhovski, V. V. Kaminski, M. S. Shur, Q. C. Chen, M. A. Khan: *Appl. Phys. Lett.* **68**, 818 (1996)
- 33.46 P. Lugli, D. K. Ferry: *IEEE Trans. Electron Dev.* **32**, 2431 (1985)
- 33.47 K. Seeger: *Semiconductor Physics: An Introduction*, 9th edn. (Springer, Berlin, Heidelberg 2004)
- 33.48 S. K. O'Leary, B. E. Foutz, M. S. Shur, L. F. Eastman: *J. Mater. Sci.: Mater. Electron.* **17**, 87 (2006)
- 33.49 W. R. L. Lambrecht, B. Segall: In: *Properties of Group III Nitrides*, EMIS Datareviews Series, ed. by J. H. Edgar (Inspec, London 1994) Chap. 4
- 33.50 M. A. Littlejohn, J. R. Hauser, T. H. Glisson: *J. Appl. Phys.* **48**, 4587 (1977)
- 33.51 J. S. Blakemore: *J. Appl. Phys.* **53**, 123 (1982)
- 33.52 T. L. Tansley, C. P. Foley: *J. Appl. Phys.* **59**, 3241 (1986)
- 33.53 V. Y. Davydov, A. A. Klochikhin, V. V. Emtsev, S. V. Ivanov, V. V. Vekshin, F. Bechstedt, J. Furthmüller, H. Harima, A. V. Mudryi, A. Hashimoto, A. Yamamoto, J. Aderhold, J. Graul, E. E. Haller: *Phys. Status Solidi B* **230**, R4 (2002)
- 33.54 J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager III., E. E. Haller, H. Lu, W. J. Schaff, Y. Saito, Y. Nanishi: *Appl. Phys. Lett.* **80**, 3967 (2002)
- 33.55 T. Matsuoka, H. Okamoto, M. Nakao, H. Harima, E. Kurimoto: *Appl. Phys. Lett.* **81**, 1246 (2002)
- 33.56 E. G. Brazel, M. A. Chin, V. Narayanamurti, D. Kapolnek, E. J. Tarsa, S. P. DenBaars: *Appl. Phys. Lett.* **70**, 330 (1997)
- 33.57 J. G. Ruch: *IEEE Trans. Electron Dev.* **19**, 652 (1972)
- 33.58 M. S. Shur, L. F. Eastman: *IEEE Trans. Electron Dev.* **26**, 1677 (1979)
- 33.59 M. Heiblum, M. I. Nathan, D. C. Thomas, C. M. Knoedler: *Phys. Rev. Lett.* **55**, 2200 (1985)

- 33.60 A. Palevski, M. Heiblum, C. P. Umbach, C. M. Knoedler, A. N. Broers, R. H. Koch: *Phys. Rev. Lett.* **62**, 1776 (1989)
- 33.61 A. Palevski, C. P. Umbach, M. Heiblum: *Appl. Phys. Lett.* **55**, 1421 (1989)
- 33.62 A. Yacoby, U. Sivan, C. P. Umbach, J. M. Hong: *Phys. Rev. Lett.* **66**, 1938 (1991)
- 33.63 E. Tiede, M. Thimann, K. Senses: *Chem. Berichte* **61**, 1568 (1928)
- 33.64 W. C. Johnson, J. B. Parsons, M. C. Crew: *J. Phys. Chem.* **36**, 2561 (1932)
- 33.65 R. Juza, H. Hahn: *Z. Anorg. Allg. Chem.* **239**, 282 (1938)
- 33.66 H. P. Maruska, J. J. Tietjen: *Appl. Phys. Lett.* **15**, 327 (1969)
- 33.67 Z. C. Huang, R. Goldberg, J. C. Chen, Y. Zheng, D. B. Mott, P. Shu: *Appl. Phys. Lett.* **67**, 2825 (1995)
- 33.68 L. F. Eastman, V. Tilak, J. Smart, B. M. Green, E. M. Chumbes, R. Dimitrov, H. Kim, O. S. Ambacher, N. Weimann, T. Prunty, M. Murphy, W. J. Schaff, J. R. Shealy: *IEEE Trans. Electron Dev.* **48**, 479 (2001)
- 33.69 C. H. Oxley, M. J. Uren: *IEEE Trans. Electron Dev.* **52**, 165 (2005)
- 33.70 M. Wraback, H. Shen, J. C. Carrano, C. J. Collins, J. C. Campbell, R. D. Dupuis, M. J. Schurman, I. T. Ferguson: *Appl. Phys. Lett.* **79**, 1303 (2001)
- 33.71 A. Matulionis, J. Liberis, L. Ardaravičius, M. Ramonas, I. Matulionienė, J. Smart: *Semicond. Sci. Technol.* **17**, 9 (2002)
- 33.72 C. Bulutay, B. K. Ridley, N. A. Zakhleniuk: *Phys. Rev. B* **68**, 115205 (2003)
- 33.73 R. Brazis, R. Raguotis: *Appl. Phys. Lett.* **85**, 609 (2004)
- 33.74 A. A. P. Silva, V. A. Nascimento: *J. Lumin.* **106**, 253 (2004)
- 33.75 S. Gökden: *Physica E* **23**, 198 (2004)
- 33.76 B. K. Ridley, W. J. Schaff, L. F. Eastman: *J. Appl. Phys.* **96**, 1499 (2004)
- 33.77 A. Matulionis, J. Liberis: *IEE Proc. Circ. Dev. Syst.* **151**, 148 (2004)
- 33.78 T. V. Shubina, S. V. Ivanov, V. N. Jmerik, M. M. Glazov, A. P. Kalavarskii, M. G. Tkachman, A. Vasson, J. Leymarie, A. Kavokin, H. Amano, I. Akasaki, K. S. A. Butcher, Q. Guo, B. Monemar, P. S. Kop'ev: *Phys. Status Solidi A* **202**, 377 (2005)

34. II–IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe

Owing to their suitable band gaps and high absorption coefficients, Cd-based compounds such as CdTe and CdS are the most promising photovoltaic materials available for low-cost high-efficiency solar cells. Additionally, because of their large atomic number, Cd-based compounds such as CdTe and CdZnTe, have been applied to radiation detectors. For these reasons, preparation techniques for these materials in the polycrystalline films and bulk single crystals demanded by these devices have advanced significantly in recent decades, and practical applications have been realized in optoelectronic devices. This chapter mainly describes the application of these materials in solar cells and radiation detectors and introduces recent progress.

34.1	Background	829
34.2	Solar Cells	829
34.2.1	Basic Description of Solar Cells	829
34.2.2	Design of Cd-Based Solar Cells	830
34.2.3	Development of CdS/CdTe Solar Cells	831
34.2.4	CdZnTe Solar Cells	834
34.2.5	The Future of Cd-Based Solar Cells	834
34.3	Radiation Detectors	834
34.3.1	Basic Description of Semiconductor Radiation Detectors	835
34.3.2	CdTe and CdZnTe Radiation Detectors	835
34.3.3	Performance of CdTe and CdZnTe Detectors	836
34.3.4	Applications of CdTe and CdZnTe Detectors	839
34.4	Conclusions	840
	References	840

34.1 Background

Cd-based compounds are very important semiconductor materials in the II–VI family. The attraction of Cd-based binary and ternary compounds arises from their promis-

ing applications as solar cells, γ - and X-ray detectors etc. These devices made from Cd-based materials are being widely applied in many fields.

34.2 Solar Cells

With the development of human society, energy sources in the earth are being slowly exhausted and we are faced with a serious problem. The solar cell is one substitute for fossil fuels and is being realized throughout the world. For this reason, solar-cell technologies have been developed since work was started by *Becquerel* in 1839 [34.1]. The solar cell has now been applied to daily life, industry, agriculture, space exploration, military affairs etc.

Solar cells have many advantages. Firstly, sunlight as an energy source for power generation is not only limitless but can be used freely. Secondly, since light is

directly converted to electricity, the conversion process is clean, noise-free and not harmful to the environment, unlike a mechanical power generator. Thirdly, solar cells need little maintenance.

34.2.1 Basic Description of Solar Cells

A solar cell is a semiconductor device that directly converts light energy into electrical energy through the photovoltaic process. The basic structure of a solar cell is shown in Fig. 34.1. A typical solar cell consists of a junction formed between an n-type and a p-type

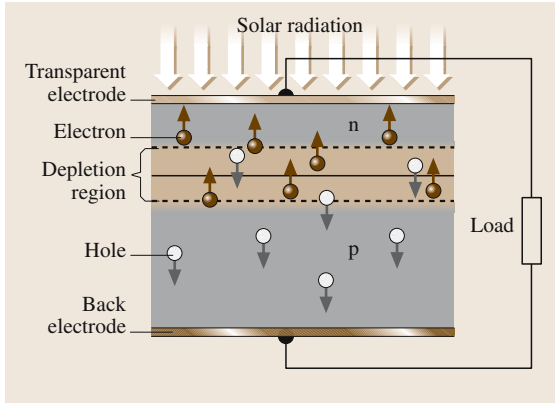


Fig. 34.1 The principle of photovoltaic devices

semiconductor, of the same (homojunction) or different materials (heterojunction), an antireflection coating and Ohmic collecting electrodes. When the light irradiates the surface of a solar cell, photons with an energy greater than the band gap of the semiconductor are absorbed by the semiconductor material. This absorption activates electron transitions from the valence to the conduction band, so that electron–hole pairs are generated. If these carriers can diffuse into the depletion region before they recombine, they can be separated by the applied electric field. At the p–n junction, the negative electrons diffuse into the n-type region and the positive holes diffuse into the p-type region. They are then collected by electrodes, resulting in a voltage difference between the two electrodes. When an external load is connected, electric current flows through the load. This is the origin of the solar cell's photocurrent.

Three parameters can be used to describe the performance of a solar cell: the short-circuit current density (J_{SC}), the open-circuit voltage (V_{OC}) and the fill factor (FF).

The short-circuit current density J_{SC} is the photocurrent output from a solar cell when the output terminals are short-circuited. In the ideal case, this is equal to the current density generated by the light J_L and is proportional to the incident photon flux. J_{SC} is determined by the spectral response of the device, the junction depth, and the series (internal) resistance, R_s . The open-circuit voltage V_{OC} is the voltage appearing across the output terminals of the cell when there is no load present, i.e. when $J = 0$. The relationship between V_{OC} and J_{SC} can be described by the diode equation, i.e.:

$$V_{OC} = \frac{nkT}{e} \ln \left(\frac{J_{SC}}{J_0} + 1 \right), \quad (34.1)$$

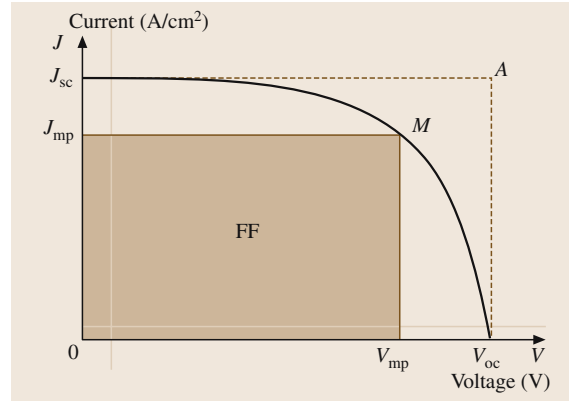


Fig. 34.2 Photovoltaic output characteristics of a solar cell

where e , n , J_0 , k and T are the charge on an electron, the diode ideality factor, the reverse saturation-current density, the Boltzmann constant and the absolute temperature, respectively. The fill factor (FF) is the ratio of the maximum electrical power available from the cell [i.e. at the operating point J_m , V_m] in Fig. 34.2 to the product of V_{OC} and J_{SC} . It describes the rectangularity of the photovoltaic output characteristic.

$$FF = \frac{V_m \cdot J_m}{V_{oc} \cdot J_{sc}} = \frac{P_m}{V_{oc} \cdot J_{sc}}. \quad (34.2)$$

The conversion efficiency (η) of a solar cell is determined as

$$\eta = \frac{P_m}{P_s} = \frac{V_{oc} \cdot J_{sc} \cdot FF}{P_s}. \quad (34.3)$$

Where P_s is the incident illumination power and P_m is the output electricity power, both being per unit area. The relationship between the short-circuit current density J_{SC} and the open-circuit voltage V_{OC} is shown in Fig. 34.2.

34.2.2 Design of Cd-Based Solar Cells

Solar cells can be made from silicon (Si), III–V compounds such as GaAs, or II–VI compounds such as Zn-based and Cd-based compound semiconductors. Of these materials, CdTe is the most attractive because of a number of advantages. CdTe is a direct-band-gap material with a band gap of 1.54 eV at room temperature. This is very close to the theoretically calculated optimum value for solar cells. CdTe has a high absorption coefficient (above 10^5 cm^{-1} at a wavelength of 700 nm), so that approximately 90% of the incident light is absorbed by a layer thickness of only $2 \mu\text{m}$ (compared

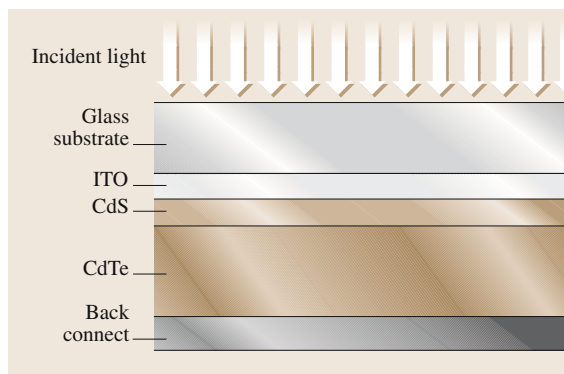


Fig. 34.3 A typical CdS/CdTe solar cell

with around $10\ \mu\text{m}$ for Si), cutting down the quantity of semiconductor required. Therefore, CdTe is one of the most promising photovoltaic materials available for use in low-cost high-efficiency solar cells.

CdTe is the only material in which both n- and p-type conductivity can be easily controlled by doping acceptor or donor impurities in II–VI compound semiconductors. Furthermore, due to its high absorption coefficient and small carrier diffusion length, the junction must be formed close to the surface, which reduces the carrier lifetime through surface recombination. In spite of this, the conversion efficiency has been gradually enhanced from 4% [34.2] to 16% [34.3] since the first CdTe-based solar cell was fabricated. CdTe solar cells can be made from polycrystalline or single-crystal material. However, it was found that this p–n homojunction structure is unstable due to the aging behavior of dopants in CdTe films and bulk single crystals [34.4]. Furthermore, it is also difficult to fabricate shallow p–n junctions with highly conducting surface layers, and high surface recombination velocities led to substantial losses that could not easily be avoided. For this reason, the CdS/CdTe heterojunction solar cell was proposed.

It has been proved that CdS/CdTe is a high-efficiency heterojunction solar cell. The theoretical calculation shows that this cell has a conversion efficiency of as high as 29% [34.5], and it is considered a promising alternative to the more widely used silicon devices. The typical structure of a CdS/CdTe solar cell is glass/indium tin oxide (ITO)/CdS/CdTe/back electrode, as seen in Fig. 34.3. The CdS/CdTe solar cell is based on the heterojunction formed between n-type CdS (only n-CdS is available) and p-type CdTe and is fabricated in a superstrate configuration where the incident light passes through the glass substrate, which has a thickness of

2–4 mm. This window glass is transparent, strong and cheap. This glass substrate also protects the active layers from the environment, and it provides the mechanical strength of the device. The outer face of the panel often has an antireflection coating to enhance absorption efficiency. A transparent conducting oxide (TCO) layer, usually tin oxide or indium tin oxide (ITO), acts as the front contact to the device. This is needed to reduce the series resistance of the device, which would otherwise arise from the thinness of the CdS layer. Usually, the polycrystalline n-type CdS layer is deposited onto the TCO. Due to its wide band gap ($E_g \approx 2.4\ \text{eV}$ at 300 K) it is transparent down to wavelengths of around 516 nm. The thickness of this layer is typically 100–300 nm. The CdTe layer is p-type-doped and its thickness is typically around $10\ \mu\text{m}$. Generally, the carrier concentration in CdS layer is 2–3 orders greater than that of the CdTe layer. As a result, the potential is applied mostly to the CdTe absorber layer, i.e. the depletion region is mostly within the CdTe layer. Therefore, the photogenerated carriers can be effectively separated in this active region.

The electrode preparation is very important in obtaining Ohmic contact or minimal series resistivity. Since there are no metals with a work function higher than the hole affinity of p-type CdTe ($-5.78\ \text{eV}$) [34.6], it is difficult to produce a complete Ohmic contact on p-type CdTe surface. Many methods have been tried to fabricate Ohmic contact to p-type CdTe. Different metals or alloy and compounds such as Au, Al, and ZnTe/Cu etc. have also been used to serve as electrode materials. Although significant effort has been applied to this issue, the junction still inevitably displays some Schottky-diode characteristics. This is a problem that will be continuously studied.

34.2.3 Development of CdS/CdTe Solar Cells

The layers of CdS and CdTe for solar cells can be prepared using various techniques, which are summarized in Table 34.1. In these techniques, evaporation, close-space sublimation (CSS), screen printing and electrodeposition (ED) have been demonstrated to be very effective for fabricating high-efficiency large-area CdS/CdTe solar cells.

Study of solar cells based on CdTe single crystals and polycrystalline films were started as early as 1963 [34.7]. The cells were prepared with a p-type copper-telluride surface film as an integral part of the photovoltaic junction. The junction was thought to be a heterojunction between the p-type copper telluride and

Table 34.1 Some fabrication techniques and conversion efficiency of CdS/CdTe solar cells

Fabrication techniques of CdS and CdTe	Conversion efficiency	area	Reference
Evaporation: physical vapor deposition (PVD), and chemical vapor deposition (CVD)	11.8%	0.3 cm ²	[34.8]
Close-spaced sublimation (CSS)	16%	1.0 cm ²	[34.3]
	8.4%	7200 cm ²	[34.9]
Electrodeposition (ED)	14.2%	0.02 cm ²	[34.10]
Screen printing (SP) and sintering	12.8%	0.78 cm ²	[34.11]
Metalorganic CVD (MOCVD)	11.9%	0.08 cm ²	[34.12]
PVD vacuum evaporation	11.8%	0.3 cm ²	[34.8]
MBE	10.5%	0.08 cm ²	[34.13]
Sputtering deposition (SD)	14%	0.09 cm ²	[34.14]

the n-type CdTe, with the transition region extending well into the CdTe side. A similar structure was reported for single-crystal cells. A conversion efficiency of up to 6% was obtained in the film cells, and 7.5% for single-crystal cells.

Dutton and Muller [34.15] fabricated CdS/CdTe solar cells by standard vapor-deposition procedures with a very thin CdTe layer. The existence of the interfacial layer was confirmed using X-ray diffraction studies, and substantiated by diode photocurrent measurements. The results showed a threshold photon energy that corresponds to the band gap of CdTe. The thin CdTe layer resulted in excellent rectification ratios and high reverse breakdown voltages. The observed photoresponse, *I*–*V* (current versus voltage), and *C*–*V* (capacitance versus voltage) characteristics were consistence with those predicted by a semiconductor heterojunction model of the CdS–CdTe interface. Mitchell et al. [34.16] prepared a variety of CdS/CdTe heterojunction solar cells with an ITO coating and a glycerol antireflection coating by vacuum evaporation of n-CdS films onto single-crystal p-CdTe substrates. Comparisons were made between cells prepared using different substrate resistivity, substrate surface preparations, and CdS film resistivity. The mechanisms of controlling the dark junction current, photocarrier collection, and photovoltaic properties were modeled, taking account of the interface states of the junction. A conversion efficiency of 7.9% was obtained under 85 mW/cm² of solar simulator illumination.

At the end of the 1970s, a new process, screen printing and sintering, was developed for fabricating CdS/CdTe polycrystal solar cells by Matsushita and coworkers [34.11, 17–20]. This technique proved to be a very effective method for fabricating large-area CdS/CdTe film solar cells. There are several important parameters in the screen-printing technique: the viscos-

ity of the paste, the mesh number of the screen, the snap-off distance between the screen and the substrate and the pressure and speed of the squeegee. As cadmium chloride (CdCl₂) has a low melting point (*T*_m = 568 °C), and forms a eutectic with both CdS and CdTe at low temperature, and the vapor pressure at 600 °C is high enough to allow complete volatilization of CdCl₂ after the sintering process, CdCl₂ is used as an ideal flux for sintering both CdS and CdTe.

Nakayama et al. [34.18] prepared the first thin-film CdS/CdTe solar cells using the screen-printing technique (Fig. 34.4). A glass plate, which was coated successively with In₂O₃ film and CdS ceramic thin film, was used as a transparent Ohmic contact substrate. The n-type CdS film with a thickness of about 20 μm and a resistivity of 0.2 Ω cm was prepared on this substrate by the screen-printing method. The CdTe paste was printed onto the CdS layer using the same technique and then heat-treated in an N₂ atmosphere at temperatures of 500–800 °C. As a result, the p-type CdTe layer with a thickness of 10 μm showed a resistivity of 0.1–1 Ωcm. A silver (Ag) paste electrode was applied to the p-type

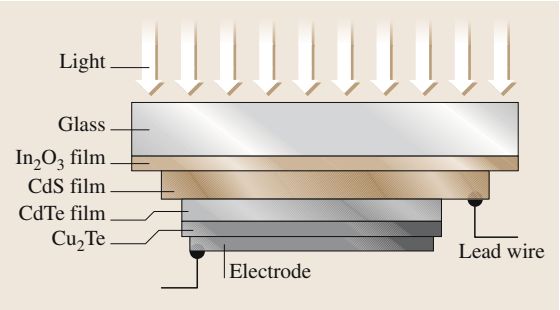


Fig. 34.4 Cross section of a ceramic thin-film CdTe solar cell

layer and an In–Ga alloy was applied to the CdS layer. The cell, with an active area of 0.36 cm^2 , showed an intrinsic solar conversion efficiency of 8.1% under an illumination of 140 mW/cm^2 solar simulator (AM0). Subsequently, this group continued to work on improving CdS/CdTe solar cells [34.11, 19, 20]. From the practical point of view, a modular solar cell is necessary. For this purpose, thin-film CdS/CdTe solar cells with an efficiency of 6.3% were prepared on a borosilicate-glass substrate of $4\times 4\text{ cm}^2$ by successively screen printing and heating (sintering) of each paste of CdS, CdTe and carbon [34.19]. Although they fabricated some structures, including dividing the CdTe film on one substrate into five small cells, to prevent increases in the series resistivity, a 1-watt module only showed an efficiency of 2.9% from 25 elemental cells with a $4\times 4\text{ cm}^2$ substrate. A main reason for this low efficiency was thought to be the high series resistivity.

In order to decrease the resistivity of the carbon electrode further, this group changed the heating conditions of the carbon electrodes for CdTe and examined the effect of impurities in the carbon paste on the characteristics of solar cells [34.11]. They found that the series resistance (R_s) and the conversion efficiency (η) of solar cells exhibited strong dependence on the oxygen (O_2) partial pressure during the heating process. The results showed that the addition of Cu to the carbon paste causes R_s and the diode factor (n) to decrease, resulting in a remarkable improvement in η . Using a low-resistance contact electrode, which was made from 50 ppm added Cu carbon paste, a cell with an active area of 0.78 cm^2 displayed $V_{oc} = 0.754\text{ V}$, $I_{sc} = 0.022\text{ A}$, $\text{FF} = 0.606$ and $\eta = 12.8\%$.

Britt and Ferekides [34.21] reported the fabrication and characteristics of high-efficiency thin-film CdS/CdTe heterojunction solar cells. CdS films with $0.07\text{--}0.10\text{ }\mu\text{m}$ in thickness were deposited on a $0.5\text{-}\mu\text{m}$

$\text{SnO}_2\text{:F}$ transparent conducting oxide layer by chemical-bath deposition and p-CdTe films with a thickness of $5\text{ }\mu\text{m}$ were deposited on the CdS layer by using the CSS method. Prior to the deposition of CdTe, CdS/ $\text{SnO}_2\text{:F}$ /glass was annealed in a H_2 atmosphere at temperatures of $350\text{--}425\text{ }^\circ\text{C}$ for 5–20 min. The characteristics of the CdS/CdTe solar cell were $V_{oc} = 843\text{ mV}$, $J_{sc} = 25.1\text{ mA/cm}^2$, $\text{FF} = 74.5\%$, corresponding to a total area greater than 1 cm^2 with an air mass 1.5 (AM1.5) conversion efficiency of 15.8%.

With modern growth techniques, high-efficiency CdS/CdTe solar cells have been developed using ultra-thin CdS films having a thickness of 50 nm [34.22]. Figure 34.5 shows the cross-sectional structure of a CdS/CdTe solar cell. CdS films were deposited on an ITO glass substrates by the metal organic chemical vapor deposition (MOCVD) technique, and CdTe films were subsequently deposited by the CSS technique (see Fig. 34.6) for the fabrication of CdS/CdTe solar cells. A thin-film CdS/CdTe solar cell with an area of 1.0 cm^2 under AM1.5 conditions showed $V_{oc} = 840\text{ mV}$, $J_{sc} = 26.08\text{ mA/cm}^2$ and $\text{FF} = 73\%$. As a result, a high conversion efficiency of 16.0% was achieved. Furthermore, they made developments in improving the uniformity of thickness and film qualities of CdS in order to obtain more efficient solar cells and to realize film deposition on large-area substrates of $30\times 60\text{ cm}$ or more [34.23]. They found clear differences in the electrical and optical properties between high- and low-quality CdS films. A photovoltaic conversion efficiency of 10.5% was achieved by a solar module with an area of 1376 cm^2 under AM1.5 measurement conditions.

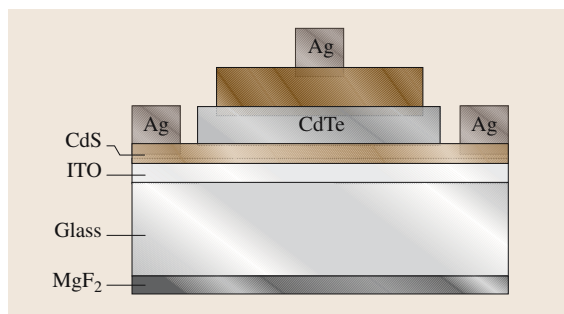


Fig. 34.5 The cross-sectional structure of a CdS/CdTe solar cell fabricating by the screen-printing method

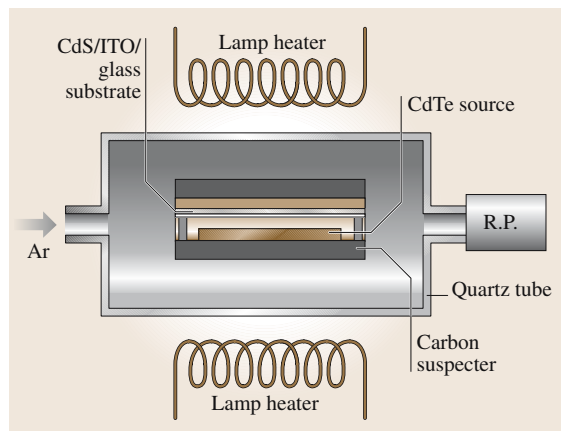


Fig. 34.6 Schematic of the close-spaced sublimation apparatus used for depositing CdTe films

In addition to dry processes such as screen printing, close-spaced sublimation etc., electrodeposition [34.24–27] has been also investigated for preparing polycrystalline CdTe layers for solar cells. *Awakura* and coworkers [34.25–27] reported the cathodic deposition behavior of CdTe thin layers under irradiation by visible light using ammoniacal basic aqueous solution as an electrolytic bath. Both deposition current density and current efficiency for the CdTe deposition were enhanced by irradiation. The deposited rate was over 10 times high than non-photoassisted electrodeposition. This was believed to represent significant progress in reducing the cost of Cd-based solar cells.

To date, solar cells with an n-CdS/p-CdTe heterojunction have been reported with efficiencies as high as 16% [34.22]. Recently, it was reported that thin-layered n-CdS/p-CdTe heterojunction solar cells have already been manufactured industrially [34.28]. n-CdS film with a thickness range of 500–1000 Å was deposited from an aqueous solution directly onto a transparent conductive oxide (TCO) substrate. After the CdS film was annealed for densification and grain growing, p-CdTe layer was formed by electrochemical deposition. The CdTe film was then annealed in air at 450 °C. Subsequently, monolithic TCO/CdS/CdTe was cut into discrete cells using an infrared laser. After other preparing process, a maximum efficiency of 10.6% for a CdTe 0.94-m² module with a power of 91.5 W was fabricated. The test results showed good stability. For practical applications, a 10-MW CdTe solar-cell manufacturing plant has been constructed [34.28].

34.2.4 CdZnTe Solar Cells

The ternary compound cadmium zinc telluride (CdZnTe, CZT) has potential for the preparation of high-efficiency tandem solar cell since its band gap can be tuned from 1.45 to 2.26 eV [34.29, 30]. *McCandless* et al. [34.29] deposited Cd_{1-x}Zn_xTe films using the physical vapor deposition (PVD) and vapor transport deposition (VTD) techniques. The film composition was between 0.35 and 0.6, corresponding to a band gap from 1.7 to 1.9 eV. Post-deposition treatment of CdZnTe films in ZnCl₂ vapor at 400 °C resulted in no change to

the alloy composition and caused recrystallization. Solar cells made from Cd_{1-x}Zn_xTe films with $x \approx 0.35$ exhibited $V_{oc} = 0.78$ V and $J_{sc} < 10$ mA/cm². These results were similar to those obtained from CdS window layers. Analysis of the spectral response indicated that Cd_{1-x}Zn_xTe with $x \approx 0.35$ has a band gap of about 1.7 eV. *Gidiputti* et al. [34.30] used two deposition technologies, co-sputtering from CdTe and ZnTe targets and co-close-spaced sublimation (CCSS) from CdTe and ZnTe powders, to prepare CdZnTe films. A structure similar to the CdTe superstrate configuration was initially utilized for cell fabrication: glass/ITO/CdS/CZT/graphite. Typical solar cell parameters obtained for CdZnTe/CdS (when $E_g(\text{CdZnTe}) = 1.72$ eV) solar cells were $V_{oc} = 720$ mV, and $J_{sc} = 2$ mA/cm². However, the spectral response indicated increasing loss of photocurrent at longer wavelengths. In order to improve the collection efficiency, CdZnTe devices were annealed in a H₂ atmosphere. This postprocessing treatment showed that J_{sc} increased to over 10 mA/cm². Study of CdZnTe solar cells is being carried out in various directions.

34.2.5 The Future of Cd-Based Solar Cells

From the development process of Cd-based solar cells, the study on the laboratory scale is being transferred to large-scale deposition and cell fabrication. Modules are being developed throughout the world, using the screen printing, evaporative deposition and close-space sublimation [34.31–33].

In order to make a commercial Cd-based photovoltaic cell with its full potential, a large-scale high-throughput manufacturing process is required. The process must possess excellent yields and produce high-efficiency devices with good long-term stability. In order to progress towards these goals, a pilot system for continuous, inline processing of CdS/CdTe devices has been developed [34.34–36]. High-quality low-cost thin-film CdTe modules with an average total area efficiency of 8% and cascaded production-line yield of > 70% have been manufactured, and technology-development programs will further increase production-line module efficiency to 13% within five years [34.36].

34.3 Radiation Detectors

A radiation detector is a device that converts a radiation ray into electrical signal. They can be divided into gas-filled detectors, scintillation detectors and

semiconductor detectors. Since semiconductor radiation detectors have a high spectrometric performance, and can be made portable, they have been applied

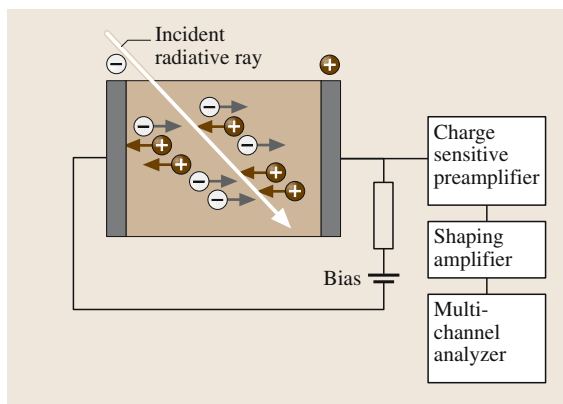


Fig. 34.7 The principle of operation of a semiconductor detector

in nuclear safeguards, medical physics and imaging diagnosis, industrial safety, nondestructive analysis, security and monitoring, nonproliferation and astrophysics [34.37, 38]. Meanwhile, they are also being gradually improved.

34.3.1 Basic Description of Semiconductor Radiation Detectors

The principle of operation of a semiconductor radiation detector is shown in Fig. 34.7. When high-energy photons (X- or γ -rays) radiate the detector, they lose the energy to induce electron–hole pairs in the semiconductor through photoelectric or Compton interactions, thereby increasing the conductivity of the material. In order to detect the change in conductivity, a bias voltage is required. An externally applied electric field separates the electron–hole pairs before they recombine, and electrons drift towards the anode, holes to the cathode. The charges are then collected by the electrodes. The collected charges produce a current pulse, whose integral equals the total charge generated by the incident photons, indicating the radiation intensity. The read-out goes through a charge-sensitive preamplifier, followed

by a shaping amplifier. The multichannel analyzer is used for spectroscopic measurements of the radiation ray. This type of detector can produce both count and energy information. Typically, detectors are manufactured in a square or rectangular configuration to maintain a uniform bias-current distribution throughout the active region. The wavelength of the peak response depends on the material's band-gap energy.

There are several important parameters to describe a detector performance. Detection efficiency is defined as the percentage of radiation incident on a detector system that is actually detected. Detector efficiency depends on the detector size and shape (larger areas and volumes are more sensitive), radiation type, material density (atomic number), and the depth of the detection medium. Energy resolution is one of the most important characteristics of semiconductor detectors. It is usually defined as the full-width at half-maximum (FWHM) of a single-energy peak at a specific energy. High resolution enables the system to more clearly resolve the peaks within a spectrum. So the reliability of the detector is also high. Detector sensitivity implies minimal detectable counts. It is defined as the number of counts that can be distinguished from the background. The relative detector response factor expresses the sensitivity of a detector relative to a standard substance. If the relative detector response factor is expressed on an equal-mass (-weight) basis, the determined sensitivity values can be substituted for the peak area.

34.3.2 CdTe and CdZnTe Radiation Detectors

Semiconductor radiation detectors are fabricated from a variety of materials including: germanium (Ge), silicon (Si), mercuric iodide (HgI_2), CdTe, CdZnTe etc.. Typical detectors for a given application depend on several factors. Table 34.2 shows the features of some semiconductor detectors. From Table 34.2, Ge detectors have the best resolution, but require liquid-nitrogen cooling, which makes them impractical for portable applications.

Table 34.2 The features of some semiconductor detectors

	NaI	Si	Ge	CdTe	$\text{Cd}_{0.9}\text{Zn}_{0.2}\text{Te}$
Bandgap (eV) at 300 K	–	1.14	0.67	1.54	1.58
Density (g/cm^3)	3.67	2.33	5.35	6.2	5.78
Energy per e–h pair (eV)	–	3.61	2.96	4.43	4.64
Detection/absorption	High	Low	Medium	High	High
Operation at room temp.	Yes	Yes	No	Yes	Yes
Spectrometric performance	Medium	High	Very high	High	High

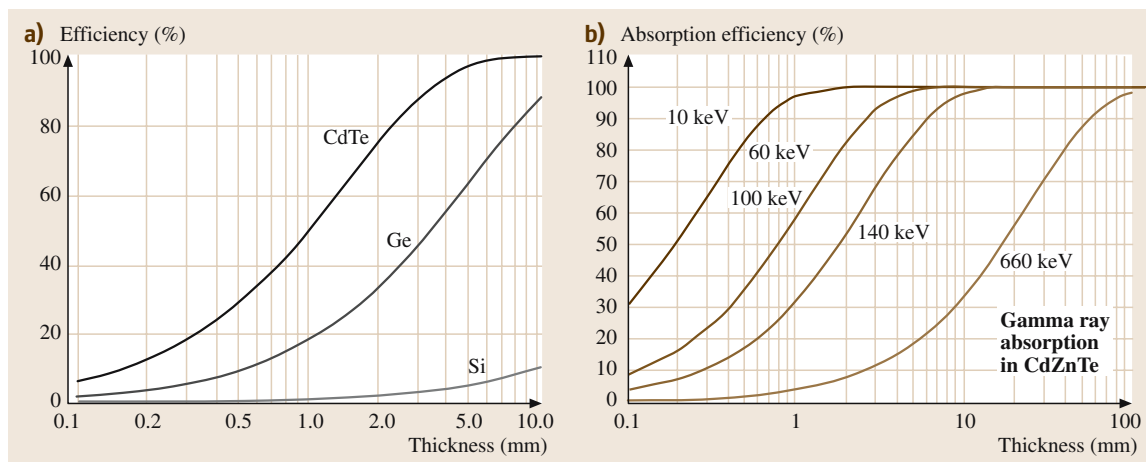


Fig. 34.8a,b γ -ray absorption in CdTe (a) and CdZnTe (b) detectors as a function of thickness

Si detectors also need cooling, and are inefficient in detecting photons with energies greater than a few tens of keV. CdTe and CdZnTe detectors possess their own advantages.

Single crystals of CdTe and CdZnTe are very important materials for the development of X- and γ -ray detectors. Currently Ge and Si detectors have to be used at liquid-nitrogen temperatures, and suffer from poor detection efficiency. Since the compound semiconductors CdTe and CdZnTe with high atomic number ($Z_{Cd} = 48$, $Z_{Zn} = 30$, $Z_{Te} = 52$) have a significantly higher photoelectric absorption efficiency in the -100–500 keV range, they can be fabricated into detectors that provide two advantages for use in portable instrumentation: (1) the large band-gap energy ($E_g > 1.54$ eV at room temperature) results in extremely small leakage current of a few nanoamps at room temperature. Therefore, these detectors have the potential to serve as useful detectors for radiation spectroscopy without the need for cryogenic cooling as required for more conventional silicon and hyper-pure germanium detectors; (2) the high density of the crystal provides excellent stopping power over a wide range of energies. The ability to operate at room temperature without the need for liquid-nitrogen cooling allows the construction of compact devices. CdTe and CdZnTe detectors can be fabricated into a variety of shapes and sizes, which makes it possible to produce detectors capable of meeting the requirements of a wide assortment of applications that are unsupported by other detector types. Their small size and relatively simple electronics allow them to open new areas of detector application. In many instances, CdTe and CdZnTe detectors

can be substituted for other detector types in existing applications.

Figure 34.8 shows the γ -ray absorption efficiency in: (a) CdTe, and (b) CdZnTe detectors as a function of thickness. It was reported that, for a detector with a thickness of about 10 mm detecting 100-keV γ -rays, the detection efficiency for a CdTe detector is 100%, for a Ge detector 88%, for a Si detector only 12% [34.39]. In the case of a CdZnTe detector, CdZnTe with a thickness of only about 7 mm can absorb 100% of the γ -rays [34.40]. This efficiency assumes that the γ -ray is completely absorbed and the generated electron-hole pairs are completely collected. In fact, a large amount of charge loss occurs in CdTe and CdZnTe detector, depending on the quality of the materials.

34.3.3 Performance of CdTe and CdZnTe Detectors

Although CdTe and CdZnTe have many advantages over Si and Ge, including higher attenuation coefficients and lower leakage currents due to the wider bandgap, their disadvantages limit their application range [34.41]. The main disadvantage is that their crystals usually contain higher density defects. These defects are remaining impurities and structural defects such as mechanical cracks, twins, grain and tilt boundaries, and dislocations all lower the crystalline perfection of the materials.

In CdTe and CdZnTe detectors, the charge transport determines the collection efficiency of generated electron-hole pairs, and structural properties of materials control the uniformity of the charge transport. The

problem of poor transport and collection of carriers, especially hole, can be overcome by applying high voltage to device and a range of electrode config. The former is related to high resistivity material, and later is related to the design of a detector. In addition, it was found that detected pulse signal intensity degraded with the time [34.42, 43]. This phenomenon was thought to be mainly related to the defects existing in bandgap. When the detector starts to work, these defects could act as slowly ionized acceptors. The low resistivity and poor quality material always exhibits such a polarization effect.

Summarizing above, improvement of a detector performance is equivalent to the improvement of the crystallinity of CdTe and CdZnTe. The key feature of all applications except substrate materials of CdTe and CdZnTe is the resistivity. This is because high resistivity can be obtained only by controlling native defects and impurity concentration. For many years, much effort was done in preparing high quality CdTe and CdZnTe single crystals.

Progresses in Crystal Growth of High Quality CdTe and CdZnTe

Various techniques have been applied to grow high-quality CdTe and CdZnTe single crystals. The Bridgman method [34.44, 45], the traveling-heater method [34.46, 47], growth from Te solvent [34.48], the gradient-freeze method [34.49, 50], and physical vapor transport [34.51, 52] are the most widely used. Crystals with applicable quality and size have become available, and have also fostered the rapid progress of research on CdTe.

As-grown CdTe single crystals commonly contain high concentration of both residual impurities and intrinsic defects. Preparation of high-purity high-resistivity CdTe and CdZnTe crystals has been widely attempted [34.53]. Early works was done by *Triboulet* and *Marfaing* [34.54] in obtaining high-purity lightly compensated zone-melting growth following synthesis by the Bridgman method. The room-temperature carrier concentrations range from 1 to $5 \times 10^{13} \text{ cm}^{-3}$, and the resistivity from 100 to $400 \Omega \text{ cm}$. The carrier mobility at 32 K reaches as high as $1.46 \times 10^5 \text{ cm}^2/\text{Vs}$. The total concentration of electrically active centers was estimated to be about 10^{14} cm^{-3} .

Usually, as-grown CdTe crystal shows p-type conductivity owing to the existence of remaining acceptor impurities or native defects. For this reason, many results have been reported in preparing high-resistivity CdTe. Chlorine (Cl) is thought to be a suitable donor for the compensation of these remaining acceptors because

Cl can be doped quite uniformly due to its very small segregation coefficient [34.55].

The growth of high-resistivity CdTe:Cl single crystals with device quality was successfully performed by the traveling-heater method (THM) [34.55]. Solvent alloys for THM growth were synthesized in ampoules filled with Te, CdTe and CdCl_2 so that the molar ratio of Te/Cd was the same as in the solvent zone during the growth. The Cl concentration in the grown crystal was 2 weight – ppm. According to the dynamics of the crystal growth, it is important to control the shape of the solid–liquid to grow high-quality single crystals. Therefore, the solvent volume was optimized. The use of a slightly tilted seed from $\langle 111 \rangle \text{B}$ was also effective in limiting the generation of twins with different directions. Single-crystal (111) wafers, larger than $30 \times 30 \text{ mm}^2$ were successfully obtained from a grown crystal with a diameter of 50 mm. Pt/CdTe/In detectors with dimensions of $2 \times 2 \times 0.5 \text{ mm}^3$ showed better energy resolution, because a higher electric field can be applied. The effective detector resistivity was estimated to be $10^{11} \Omega \text{ cm}$.

CdTe doped with chlorine (Cl) or indium (In) with a resistivity of $3 \times 10^9 \Omega \text{ cm}$ and CdZnTe with a resistivity of $5 \times 10^{10} \Omega \text{ cm}$ were grown by the high-pressure Bridgman (HPB) technique [34.56]. The material was polycrystalline with large grains and twins. Although the crystalline quality of HPB CdTe and CdZnTe is poor, the grains are large enough to obtain volume detectors of several cm^3 . Photoluminescence (PL) spectra at 4 K showed that the free exciton could be observed and the FWHM of the bound excitons is very low. Some impurity emissions were identified. The detectors were fabricated from HPB CdTe and CdZnTe. These detectors showed excellent performance. Gamma-ray spectra were presented with high-energy resolution in an energy range from 60 to 600 keV. Using a $10 \times 10 \times 2 \text{ mm}^2$ HPB detector, at a bias of 300 V, the peak at 122 keV from ^{57}Co had a FWHM of 5.2 KeV. Detectors with high-energy resolution were fabricated.

Because of the difference in vapor pressure between Cd and Te, grown crystals contain a large number of Cd vacancies (V_{Cd}). These Cd vacancies manifest acceptor behaviour. Therefore, high-resistivity CdTe can also be obtained by controlling the concentration of Cd vacancies. However, a prerequisite is that the purity of the CdTe crystal is high enough that the remaining impurities do not play a substantial role in determining the conductivity. Recently, the preparation of ultra-high-purity CdTe single crystals was reported [34.57, 58]. In order to obtain a high purity the starting mater-

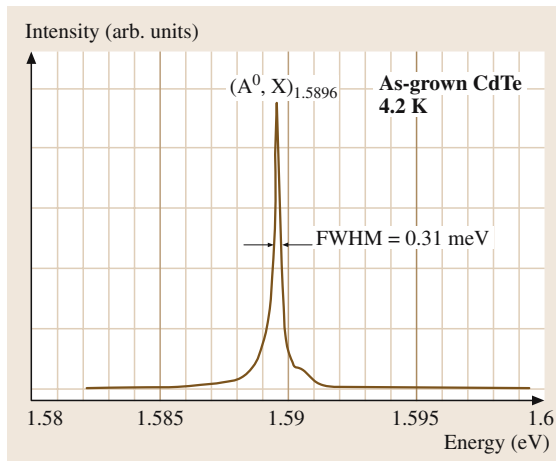


Fig. 34.9 A typical high-resolution PL spectrum of a high-purity CdTe sample

ials of Cd and Te, Cd were first purified by vacuum distillation (VD) and the overlap zone-melting (OZM) method [34.57], while Te was purified using the normal freezing method [34.58]. The results of glow-discharge mass spectroscopy (GDMS) and the measurement of the residual resistivity ratio (RRR) showed that refined Cd and Te had the purity of 6N-up. Using the refined Cd and Te as starting materials, extremely high-purity twin-free CdTe single crystals were prepared by the traditional vertical Bridgman technique. Figure 34.9 shows a typical high-resolution PL spectrum of a high-purity CdTe sample. Only a sharp and strong emission line of (A^0, X) (an exciton bound to a neutral acceptor) at 1.5896 eV was observed. The FWHM of the (A^0, X) line is 0.31 meV. This is the narrowest value ever reported for CdTe single crystals grown by the Bridgman method. All of the above indicate that the sample is of high purity and quality. For a good performance detector, the leakage current must not exceed a few nanoamps, which requires a material with high resistivity. When growing CdTe crystals, the CdTe contain large number of Cd vacancies due to the extreme volatility of Cd. For these reason, CdTe single crystals were annealed in a Cd atmosphere in order to obtain a high-resistivity crystal. Hall-measurement results showed that the conductive type changed from as-grown p-type to n-type at a Cd pressure around 1.5×10^{-2} atm, corresponding to a Cd source temperature of 500 °C. Samples annealed under these conditions showed a resistivity as high as $10^9 \Omega\text{cm}$.

Lachish [34.59] grew CdTe and CdZnTe crystals with excess cadmium in order to avoid the Te precip-

itates found in crystals grown by Bridgman method. During crystal growth one ampoule end was kept at a low temperature, which determines a constant, nearly atmospheric, vapor pressure in the system. The constant vapor pressure maintains a constant liquid composition and balanced amounts of cadmium and tellurium within the crystal. Although the highly donor-doped crystals are highly electrically conducting, successive annealing in Te vapor transforms the crystal wafers into a highly compensated state showing high electrical resistivity and high gamma sensitivity, depending on the doping level.

Improvements of Transport and Collection of Carriers in CdTe and CdZnTe Detectors

The transport and collection of carriers in CdTe and CdZnTe, especially holes, are the main problems that affect their performance. Poor carriers transport causes position-dependent charge collection and limits their capability as high-resolution spectrometers [34.60]. As a result, this limits the volume of detector. Therefore, the efficient detecting resolution and range are decreased. Amman et al. [34.61] studied the affect of nonuniform electron trapping on the performance of a γ -ray detector. An analysis of the induced charge signals indicated that regions of enhanced electron trapping are associated with inclusions, and that these regions extend beyond the physical size of the inclusions. Such regions introduce nonuniform electron trapping in the material that then degrades the spectroscopic performance of the material as a γ -ray detector. The measurements showed that the degree of nonuniformity that affects detector performance could be at the 1% level. Consequently, any useful characterization and analysis technique must be sensitive down to this level.

Detectors equipped with Ohmic contacts, and a grounded guard ring around the positive contact, have a fast charge-collection time. Conductivity adjustment, in detectors insensitive to hole trapping, optimizes the detector operation by a trade-off between electron lifetime and electrical resistance. Better hole collection has been achieved by designing detectors and electrodes of various geometries. Nakazawa et al. [34.62] improved the CdTe diode detectors by fabricating the guard-ring structure in the cathode face, and the leakage current of the detector decreased by more than an order of magnitude. The new CdTe detector was operated with a bias as high as 800 V at 20 °C and showed a good energy resolution (0.93 keV and 1 keV FWHM for 59.5 keV and 122 keV, respectively) and high stability for long-term operation at room temperature. Furthermore, they also developed a large-area ($20 \times 20 \text{ mm}^2$) detector with

very high resolution. Owing to its high stopping power and high resolution at room temperature, pixel imagers have also been developed in the same group, and it is thought that this large diode detector could possibly be a substitute for scintillation detectors.

Since the carriers drift slow and have a short lifetime in CdTe detectors, the number of photons in the photo-peak is reduced and the spectrum is distorted by a tail towards lower energies. For these reasons, CdTe detectors are usually made into Schottky diodes, because this structure can withstand much higher bias voltage with a leakage current orders of magnitude lower than detectors with Ohmic contacts. *Takahashi et al.* [34.63] adopted the configuration of Schottky CdTe diode, which due to the low leaking current, makes it possible to apply a much higher bias voltage to ensure complete charge collection in relatively thin (< 1 mm) devices. Both the improved charge-collection efficiency and the low leakage current lead to an energy resolution of better than 600 eV FWHM at 60 keV for a 2×2 mm² device without any electronics for charge-loss correction. Meanwhile, they also fabricated large-area detectors with dimensions of 21.5×21.5 mm², with a thickness of 0.5 mm and an energy resolution of 2.8 keV. Stacked detectors can measure the energies as high as 300 keV. Furthermore, a large-array detector, consisting of 1024 individual CdTe diodes, was also made. Every detector had a dimension of 1.2×5.0 mm². The total area, including the spaces between the detector elements, is 44×44 mm². This array detector is expected to be used in next-generation Compton telescopes.

34.3.4 Applications of CdTe and CdZnTe Detectors

Owing to the convenience of the smaller collimator, better resolution and temperature stability, CdTe and CZT detectors have been used in safeguard applications by the International Atomic Energy Agency (IAEA) and some countries for over ten years [34.64, 65]. With the gradual improvement in performance, CdTe and CdZnTe detectors are replacing NaI detectors used in spent-fuel attribute tests [34.66], though their sensitivity is still low compared to NaI and Ge detectors, although already sufficient for many applications. CdTe detectors have a sensitive volume of about 20–100 mm³ and a probe diameter of 8–9 mm. CdZnTe detectors have a larger volume than CdTe detectors. The largest commercial CdZnTe detectors have a geometric volume of 1687 mm³ ($15 \times 15 \times 7.5$ mm³). Detectors are mainly of hemispheric design to obtain

high carrier-collection efficiency. These large-volume detectors have been made into portable and hand-held isotope-identification devices and are being used to detect radioactive sources. They will become commercially available in the near future.

At present, conventional X-ray film or scintillator mammograms are used in medical diagnostics such as screening for breast cancer. However, these show a non-linear response to X-ray intensity and the detection quantum efficiency is low. Room-temperature semiconductor detectors such as CdTe and CdZnTe have favorable physical characteristics for medical applications. From the start of these investigations in the 1980s, rapid progress has been achieved [34.67–72]. *Barber* [34.68] presented results concerning, first, a CdTe two-dimensional (2-D) imaging system (20×30 mm² with 400×600 pixels) for dental radiology and, second, a CdZnTe fast pulse-correction method applied to a $5 \times 5 \times 5$ mm³ CdZnTe detector (energy resolution of 5% for a detection efficiency of 85% at 122 keV) for medical imaging. After that, a 2-mm-thick CdZnTe detector was fabricated for application to digital mammography [34.70]. The preliminary images showed high spatial resolution and efficiency. Furthermore, CdTe and CdZnTe detectors with a thickness of 0.15–0.2 mm were fabricated [34.72]. The detectors are indium-bump-bonded onto a small version of a chip. Their detection quantum efficiency was measured as 65%. This result showed that CdTe and CdZnTe detectors are superior to scintillator-based digital systems, whose quantum efficiency is typically around 30–40%. This showed that CdTe and CdZnTe detectors have potential applications in medical imaging, as well as industrially for nondestructive evaluation inspection.

In universe exploration, Cd-based detectors can be used in an advanced Compton telescope (ACT) planned as the next-generation space-based instrument devoted to observations of low/medium-energy γ -rays (≈ 0.2 –30 MeV) and to the nonthermal energy exploration telescope (NeXT) [34.73]. In the universe, radiation rays have a wide energy range of 0.5–80 keV. In order to detect this wide range of radiation, *Takahashi et al.* [34.74–76] proposed a new focal-plane detector based on the idea of combining an X-ray charge-coupled device (CCD) and a CdTe pixel detector as the wide-band X-ray imager (WXI). The WXI consists of a soft-X-ray imager and a hard-X-ray imager. For the detection of soft X-rays (10–20 keV) with high positional resolution, a CCD with a very thin dead layer will be used. For hard X-rays, CdTe pixel detectors serve as absorbers. This study is now under way.

34.4 Conclusions

Devices fabricated from Cd-based compounds, such as solar cells and radiation detectors, are being applied in our daily life. In the case of solar cells, although recent success have improved their conversion efficiency and reduced the cost, many problems remain. Fundamental understanding of the CdTe-based solar-cell properties is limited, particularly as a result of their polycrystalline nature. Therefore, the fundamental electronic properties of polycrystalline Cd-based thin films should be studied deeply. Other challenges are to reduce the cost and to lengthen the operating life span. These problems are being studied and solved [34.77].

In the case of radiation detectors, CdTe and CdZnTe detectors have many advantages, such as room-temperature operation, high count rates, small size, and direct conversion of photons to charge, which make them attractive candidates for a wide variety of applications in industrial gauging and analysis, as well as medical instrumentation and other areas. These detectors are being made available commercially at present. However, they have severe problems such as polarization effects, long-term stability and their high price. Further efforts should still be focused on the preparation of high-quality materials and improvement of the stability and reliability of detectors. We are confident that radiation detectors made from

Cd-based compounds will achieve more widespread application.

Definition of Terms

- **Cd-based compound semiconductor:** a semiconductor that contains the element Cd.
- **Heterojunction:** a junction between semiconductors that differ in their doping-level conductivities, and also in their atomic or alloy composition.
- **Band gap:** energy difference between the conduction band and the valence band.
- **n-type conductivity:** a semiconductor material, with electrons as the majority charge carriers, that is formed by doping with donor atoms.
- **p-type conductivity:** a semiconductor material in which the dopants create holes as the majority charge carrier, formed by doping with acceptor atoms.
- **Solar cell:** a semiconductor device that converts the energy of sunlight into electric energy. Also called a photovoltaic cell.
- **Semiconductor detector:** a device that converts the incident photons directly into an electrical pulse.
- **Conversion efficiency:** the ratio of incident photon energy and output electricity energy.
- **Detection efficiency:** percentage of radiation incident on a detector system that is actually detected.

References

- 34.1 E. Becquerel: Compt. Rend. Acad. Sci. (Paris) **9**, 561 (1839)
- 34.2 Yu. A. Vodakov, G. A. Lomakina, G. P. Naumov, Yu. P. Maslakovets: Sov. Phys. Solid State **2**, 1 (1960)
- 34.3 T. Aramoto, S. Kumazawa, H. Higuchi, T. Arita, S. Shibutani, T. Nishio, J. Nakajima, M. Tsuji, A. Hanafusa, T. Hibino, K. Omura, H. Ohyama, M. Murozono: Jpn. J. Appl. Phys. **36**, 6304 (1997)
- 34.4 B. Yang, Y. Ishikikawa, T. Miki, Y. Doumae, M. Ishiki: J. Cryst. Growth **179**, 410 (1997)
- 34.5 A. W. Brinkman: Properties of Narrow Gap Cadmium-Based Compounds. In: *Electronic Materials Information Services*, Vol. 10, ed. by P. Capper (IEE, London 1994) p. 591
- 34.6 R. W. Swank: Phys. Rev. **156**, 844 (1967)
- 34.7 D. A. Cusano: Solid State Electron. **6**, 217 (1963)
- 34.8 R. G. Little, M. J. Nowlan: Progress in Photovoltaics **5**, 309 (1997)
- 34.9 Y.-S. Tyan, E. A. Perez-Albuern: In: *Proc. 16th IEEE Photovoltaic Specialists Conf.* (IEEE, New York 1982) p. 794
- 34.10 J. M. Woodcock, A. K. Turner, M. E. Özsan, J. G. Summers: In: *Proc. 22nd IEEE Photovoltaic Specialists Conf., Las Vegas* (IEEE, New York 1991) p. 842
- 34.11 K. Kuribayashi, H. Matsumoto, H. Uda, Y. Komatsu, A. Nakano, S. Ikegami: Jpn. J. Appl. Phys. **22**, 1828 (1993)
- 34.12 J. Britt, C. Ferekides: Appl. Phys. Lett. **62**, 2851 (1993)
- 34.13 H. W. Schock, A. Shah: In: *Proc. 14th European Photovoltaics Solar Energy Conf.*, ed. by H. A. Osssenbrink, P. Helm, H. Ehmann (H. S. Stephens & Ass., Bedford, UK 1997) p. 2000
- 34.14 A. D. Compaan, A. Gupta, J. Drayton, S.-H. Lee, S. Wang: Phys. Stat. Solid B **241**, 779 (2004)
- 34.15 R. W. Dutton, R. S. Muller: Solid State Electron. **11**, 749 (1968)

- 34.16 K.W. Mitchell, A.L. Fahrenbruch, R.W. Bube: *J. Appl. Phys.* **48**, 4365 (1977)
- 34.17 H. Uda, A. Nakano, K. Kuribayashi, Y. Komatsu, H. Matsumoto, S. Ikegami: *Jpn. J. Appl. Phys.* **22**, 1822 (1983)
- 34.18 N. Nakayama, H. Matsumoto, K. Yamaguchi, S. Ikegami, Y. Hioki: *Jpn. J. Appl. Phys.* **15**, 2281 (1976)
- 34.19 S. Ikegami, T. Yamashita: *J. Electron. Mater.* **8**, 705 (1979)
- 34.20 N. Nakayama, H. Matsumoto, A. Nakano, S. Ikegami, H. Uda, T. Yamashita: *Jpn. J. Appl. Phys.* **19**, 703 (1980)
- 34.21 J. Britt, C. Ferikides: *Appl. Phys. Lett.* **62**, 2851 (1993)
- 34.22 T. Aramoto, S. Kumazawa, H. Higuchi, T. Arita, S. Shibutani, T. Nishio, J. Nakajima, M. Tsuji, A. Hanafusa, T. Hibino, K. Omura, H. Ohyama, M. Murozono: *Jpn. J. Appl. Phys.* **36**, 6304 (1997)
- 34.23 M. Tsuji, T. Aramoto, H. Ohyama, T. Hibino, K. Omura: *Jpn. J. Appl. Phys.* **39**, 3902 (2000)
- 34.24 M. P. R. Panicker, M. Knaster, F. A. Kröger: *J. Electrochem. Soc.* **125**, 566 (1978)
- 34.25 K. Murase, H. Uchida, T. Hirato, Y. Awakura: *J. Electrochem. Soc.* **146**, 531 (1999)
- 34.26 K. Murase, M. Matsui, M. Miyake, T. Hirato, Y. Awakura: *J. Electrochem. Soc.* **150**, 44 (2003)
- 34.27 M. Miyake, K. Murase, H. Inui, T. Hirato, Y. Awakura: *J. Electrochem. Soc.* **151**, 168 (2004)
- 34.28 D.W. Cunningham, M. Rubcich, D. Skinner: *Prog. Photovoltaics* **10**, 59 (2002)
- 34.29 B. McCandless, K. Dobson, S. Hegedus, P. Paulson: *NCPV and Solar Program Review Meeting Proceeding, March 24–26 2003, Denver, Colorado* (NREL, Golden, Colorado 2003) p.401 Available in electronic form, NREL/CD-520-33586
- 34.30 G. Gidiputti, P. Mahawela, M. Ramalingan, G. Sivaraman, S. Subramanian, C.S. Ferekides, D.L. Morel: *NCPV and Solar Program Review Meeting Proceeding, March 24–26 2003, Denver, Colorado* (NREL, Golden, Colorado 2003) p.896 Available in electronic form, NREL/CD-520-33586
- 34.31 P. D. Maycock: *PV News* **17**, 3 (1998)
- 34.32 R. C. Powell, U. Jayamaha, G. L. Dorer, H. McMaster: *Proc. NCPV Photovoltaics Program, Review*, ed. by M. Al-Jassim, J. P. Thornton, J. M. Gee (American Institute of Physics, New York 1995) p.1456
- 34.33 D. Bonnet, H. Richter, K.-H. Jager: *Proc. 13th European Photovoltaic Solar Energy Conference*, ed. by W. Freiesleben, W. Palz, H. A. Ossenbrink, P. Helm (Stephens, Bedford, UK 1996) p.1456
- 34.34 K. Zweibel, H. Ullal: *Proceeding of the 25th IEEE Photovoltaic Specialists Conference, Washington DC 1996* (IEEE, New York 1996) p.745
- 34.35 K. L. Barth, R. A. Enzenroth, W. S. Sampath: *NCPV and solar Program Review Meeting Proceeding, March 24–26 2003, Denver, Colorado* (NREL, Golden, Colorado 2003) p.904 Available in electronic form, NREL/CD-520-33586
- 34.36 A. Abken, C. Hambro, P. Meyers, R. Powell, S. Zafar: *NCPV and solar Program Review Meeting Proceeding, March 24–26 2003, Denver, Colorado* (NREL, Golden, Colorado 2003) p.393 Available in electronic form, NREL/CD-520-33586
- 34.37 K. Zanio: , Vol. 13 (Academic, New York 1978) p.164
- 34.38 R. Triboulet, Y. Mafaing, A. Cornet, P. Siffert: *J. Appl. Phys.* **45**, 2759 (1974)
- 34.39 G. Sato, T. Takahashi, M. Sugiho, M. Kouda, T. Mitani, K. Nakazawa, Y. Okada, S. Watanabe: *IEEE Trans. Nucl. Sci.* **48**, 950 (2001)
- 34.40 C. Szeles: *Phys. Stat. Solid B* **241**, 783 (2004)
- 34.41 H. Yoon, J. M. Van Scyoc, T. S. Gilbert, M. S. Goorsky, B. A. Brunett, J. C. Lund, H. Hermon, M. Schieber, R. B. James: *Infrared Applications of Semiconductors II. Symposium, Boston, MA, USA, 1–4 Dec. 1997*, ed. by D.L. McDaniel Jr., M.O. Manasreh, R. H. Miles, S. Sivananthan, P. A. Warrendale (Materials Research Society, Pittsburgh, PA 1998) p.241. USA: Mater. Res. Soc, 1998
- 34.42 R. O. Bell, G. Entine, H. B. Serreze: *Nucl. Instrum. Methods* **117**, 267 (1974)
- 34.43 P. Siffert, J. Berger, C. Scharager, A. Cornet, R. Stuck, R. O. Bell, H. B. Serreze, F. V. Wald: *IEEE Trans. Nucl. Sci.* **23**, 159 (1976)
- 34.44 R. K. Route, M. Woff, R. S. Feigelson: *J. Cryst. Growth* **70**, 379 (1984)
- 34.45 K. Y. Lay, D. Nichols, S. McDevitt, B. E. Dean, C. J. Johnson: *J. Cryst. Growth* **86**, 118 (1989)
- 34.46 R. O. Bell, N. Hemmat, F. Wald: *Phys. Stat. Solid A* **1**, 375 (1970)
- 34.47 R. Triboulet, Y. Mafaing, A. Cornet, P. Siffert: *J. Appl. Phys.* **45**, 375 (1970)
- 34.48 K. Zanio: *J. Electron. Mat.* **3**, 327 (1974)
- 34.49 M. Azoulay, A. Raizman, G. Gafni, M. Roth: *J. Cryst. Growth* **101**, 256 (1990)
- 34.50 A. Tanaka, Y. Masa, S. Seto, T. Kawasaki: *Mater. Res. Soc. Symp. Proc.* **90**, 111 (1987)
- 34.51 W. Akutagawa, K. Zanio: *J. Cryst. Growth* **11**, 191 (1971)
- 34.52 C. Ceibel, H. Maier, R. Schmitt: *J. Cryst. Growth* **86**, 386 (1988)
- 34.53 M. Isshiki: *Wide-gap II–VI Compounds for Optoelectronic Applications* (Chapman Hall, London 1992) p.3
- 34.54 R. Triboulet, Y. Mafaing: *J. Electrochem. Soc.* **120**, 1260 (1973)
- 34.55 M. Funaki, T. Ozaki, K. Satoh, R. Ohno: *Nucl. Instr. Meth. A* **322**, 120 (1999)
- 34.56 M. Fiederle, T. Feltgen, J. Meinhardt, M. Rogalla, K. W. Benz: *J. Cryst. Growth* **197**, 635 (1999)
- 34.57 B. Yang, Y. Ishikawa, Y. Doumae, T. Miki, T. Ohyama, M. Isshiki: *J. Cryst. Growth* **172**, 370 (1997)
- 34.58 S. H. Song, J. Wang, M. Isshiki: *J. Cryst. Growth* **236**, 165 (2002)
- 34.59 <http://urila.tripod.com/crystal.htm>

- 34.60 T. Takahashi, S. Watanabe: IEEE Trans. Nucl. Sci. **48**, 950 (2001)
- 34.61 M. Amman, J. S. Lee, P. N. Luke: J. Appl. Phys. **92**, 3198 (2002)
- 34.62 K. Nakazawa, K. Oonuki, T. Tanaka, Y. Kobayashi, K. Tamura, T. Mitani, G. Sato, S. Watanabe, T. Takahashi, R. Ohno, A. Kitajima, Y. Kuroda, M. Onishi: IEEE Trans. Nucl. Sci. **51**, 1881 (2004)
- 34.63 T. Takahashi, T. Mitani, Y. Kobayashi, M. Kouda, G. Sato, S. Watanabe, K. Nakazawa, Y. Okada, M. Funaki, R. Ohno, K. Mori: IEEE Trans. Nucl. Sci. **49**, 1297 (2002)
- 34.64 R. Arlt, D. E. Rundquist: Nucl. Instr. Methods Phys. Res. A **380**, 455 (1996)
- 34.65 T. Prettyman: *2nd Workshop on Science and Modern Technology for Safeguards, Albuquerque, NM, U.S.A., 21–24 September 1998*, ed. by C. Foggi, E. Petraglia (European Commission, Albuquerque, NM 1998)
- 34.66 W. K. Yoon, Y. G. Lee, H. R. Cha, W. W. Na, S. S. Park: INMM J. Nucl. Mat. Manage. **27**, 19 (1999)
- 34.67 C. Scheiber, J. Chambron: Nucl. Instr. Meth. A **322**, 604 (1992)
- 34.68 H. B. Barber: J. Electron. Mater. **25**, 1232 (1996)
- 34.69 L. Verger, J. P. Bonnefoy, F. Glasser, P. Ouvrier-Buffet: J. Electron. Mater. **26**, 738 (1997)
- 34.70 S. Yin, T. O. Tümay, D. Maeding, J. Mainprize, G. Mawdsley, M. J. Yaffe, W. J. Hamilton: IEEE Trans. Nucl. Sci. **46**, 2093 (1999)
- 34.71 C. Scheiber: Nucl. Instr. Meth. A **448**, 513 (2000)
- 34.72 S. Yin, T. O. Tümay, D. Maeding, J. Mainprize, G. Mawdsley, M. J. Yaffe, E. E. Gordon, W. J. Hamilton: IEEE Trans. Nucl. Sci. **49**, 176 (2002)
- 34.73 T. Tanaka, T. Kobayashi, T. Mitani, K. Nakazawa, K. Oonuki, G. Sato, T. Takahashi, S. Watanabe: New Astron. Rev. **48**, 269 (2004)
- 34.74 T. Takahashi, B. Paul, K. Hirose, C. Matsumoto, R. Ohno, T. Ozaki, K. Mori, Y. Tomita: Nucl. Instr. Meth. A **436**, 111 (2000)
- 34.75 T. Takahashi, K. Nakazawa, T. Kamae, H. Tajima, Y. Fukazawa, M. Nomachi, M. Kokubun: SPIE **4851**, 1228 (2002)
- 34.76 T. Takahashi, K. Makishima, Y. Fukazawa, M. Kokubun, K. Nakazawa, M. Nomachi, H. Tajima, M. Tashiro, Y. Terada: New Astron. Rev. **48**, 309 (2004)
- 34.77 V. K. Krishna, V. Dutta: J. Appl. Phys. **96**, 3962 (2004)

35. Doping Aspects of Zn-Based Wide-Band-Gap Semiconductors

The present Chapter deals with the wide-band-gap (defined here as greater than 2 eV) Zn chalcogenides, i. e. ZnSe, ZnS, and ZnO (mainly in bulk form). However, since recent literature on ZnS is minimal, the main coverage is of ZnSe and ZnO. In addition $\text{Zn}_{1-x}\text{Be}_x\text{Se}$ ($x \leq 0.5$) is included, since Be is expected to reduce degradation (from light irradiation/emission) in ZnSe. The main emphasis for all these materials is on doping, in particular p-type doping, which has been a problem in all cases. In addition, the origin of light emission

35.1	ZnSe	843
35.1.1	Doping – Overview	843
35.1.2	Results on p-Type Material with N as the Primary Dopant	845
35.2	ZnBeSe	848
35.3	ZnO	849
35.3.1	Doping	849
35.3.2	Optical Properties	850
	References	851

in ZnO is not yet well established, so this aspect is also briefly covered.

The present Chapter treats the wide-band-gap (defined here as greater than 2 eV) Zn chalcogenides (as well as ZnBeSe), i. e., ZnSe, ZnS, and ZnO, with room-temperature band gaps of 2.7 eV, 3.7 eV, and 3.4 eV, respectively. We shall here concentrate mainly on bulk properties, since quantum dots and quantum wells are treated elsewhere in this Handbook except when these (or other nanostructures) are involved in bulk doping (Sect. 35.1.2). The primary emphasis will be on literature from 2000 to 2004. Moreover, since there have been few publications on ZnS in the last four years (our literature search showed only seven publications) [35.1–7], the present review will effectively cover ZnSe, ZnBeSe, and ZnO.

It is well known that the primary interest in these materials is their ability to provide light emission and/or detection in the green and higher spectral ranges. One of the major problems for these materials is obtaining good bipolar doping, in particular good p-type doping

for ZnO, ZnSe, and ZnBeSe with low fractions of Be; this problem has for instance been reviewed for ZnO by Pearton et al. [35.8] and by Look and Claflin [35.9] and for ZnSe by Neumark [35.10]. A second problem, especially for ZnSe-based devices, is that of degradation under photon irradiation, including those generated during light emission [35.11–14]. It is for this reason that ZnBeSe is of high interest, since Be is expected to *harden* ZnSe, i. e. to reduce defect formation and thus degradation [35.15–17].

ZnO is one of the most studied materials in the group of II–VI semiconductors because of its wide band gap (3.36 eV at room temperature) and its bulk exciton-binding energy (60 meV), which is larger than the room-temperature thermal energy. In addition to room-temperature ultraviolet (UV) optoelectronic devices, it can be used for magnetic [35.18] and biomedical applications [35.19] and references therein.

35.1 ZnSe

35.1.1 Doping – Overview

Despite many years of effort, p-type doping of ZnSe is still a problem. The main success to date has been achieved with nitrogen as the primary dopant. Of other dopants, Li diffuses extremely fast [35.20] and also self-compensates via interstitial Li [35.20, 21], Na has a predicted maximum equilibrium solubility of

$5 \times 10^{17} \text{ cm}^{-3}$ [35.22] and also self-compensates (via interstitial Na), as shown by Neumark et al. [35.23], P and As give DX centers and thus give deep levels (as summarized for instance by Neumark [35.10]), and Sb to date has given net acceptor concentrations of only about 10^{16} cm^{-3} (see Table 35.1). Regarding, N doping, Table 35.2 lists recent results on concentrations of holes (p) or net acceptors ($[n_a - n_d]$), where n_a (n_d) is the accep-

Table 35.1 p-type doping of ZnSe with dopants other than N

Dopant	p (cm ⁻³)	n _a - n _d (cm ⁻³)	E _a (meV)	Method	Reference
Sb		1.5 × 10 ¹⁶	69	MOVPE	[35.25]
Sb	≈ 10 ¹⁶			MOVPE	[35.26]
Sb			55 ± 5	MOVPE	[35.27]
Sb		(7 ± 3) × 10 ¹⁶		PVT	[35.28]
K	9 × 10 ¹⁷			Eximer laser	[35.29]
Na	5 × 10 ¹⁹			Eximer laser	[35.29]
Co-doping Li, I		2 × 10 ¹⁶		MOVPE	[35.30]
Co-doping Li, Cl		3.8 × 10 ¹⁶		MBE	[35.31]
GaAs:Zn nano-cluster		1 × 10 ¹⁷		MOMBE ^a	[35.32]

^ametalorganic molecular beam epitaxy

Table 35.2 Doped ZnSe with p or (n_a - n_d) above 10¹⁸ cm⁻³

	Best p or n _a - n _d (cm ⁻³)	E _a (meV)	Comments on degradation	Reference
Sub-monolayer (N + Te) δ ³ -doped	6 × 10 ¹⁸	38–87	Expected to be minimal	[35.24, 33]
Li ₃ N diffusion	8 × 10 ¹⁸			[35.34]
MOVPE-grown N-doped	1 × 10 ¹⁸			[35.35]
ZnSe/ZnTe:N δ-doped superlattice	7 × 10 ¹⁸	30	Expected to be high	[35.36]

Table 35.3 n-type doping of ZnSe

Dopant	n (cm ⁻³)	Method	Comments	Reference
Cl			The PL is dominated by the Cl ⁰ X line at 2.797 eV (10 K). Above 200 K, the intensity of the Cl ⁰ X line decreases rapidly due to the presence of a nonradiative center with a thermal activation energy of ≈ 90 meV. The decrease of the Cl ⁰ X line over the temperature range 10–200 K is due to the thermal activation of the Cl ⁰ X line bound exciton to a free exciton with abactivation energy of ≈ 9.0 meV	[35.37]
Cl		MBE	At high ZnCl ₂ beam intensity, crystallinity deteriorates due to excess Cl atom	[35.38]
Cl		MBE	At low T, the dominate PL is due to neutral donor-bound excitons; at high T, the dominate PL is due to free-to-bound recombination. At low T, two additional lines on the high-energy side are observed (light- and heavy-hole free-exciton transitions); one additional peak at the low-energy side (DAP ^a transition)	[35.39]
Al	4.2 × 10 ¹⁸ – 1.2 × 10 ¹⁹	MBE	Three deep levels are reported: an acceptor-like state at 0.55 eV above VBM ^b and two donor states at 0.16 eV and 0.80 eV below CBM ^c	[35.40, 41]
Br	1.4–4.1 × 10 ¹⁷	Vertical sublimation		[35.42]
Br	4.0 × 10 ¹⁶	PVT	Two deep electron traps with thermal activation energy 0.20 eV and 0.31 eV are reported	[35.43]
In		Dopant diffusion	A temperature range can be found where electron concentration decreases with an increase in temperature	[35.44]

^adonor–acceptor pair; ^bvalence band maximum; ^cconduction band maximum

tor (donor) concentration, in various approaches, where these are greater than 10^{18} cm^{-3} . We note in connection with Table 35.2 that degradation associated with N can be a severe problem [35.11, 12], and we also give some comments on degradation in the table. We shall discuss two N-doped systems in more detail below Sect. 35.1.2. One uses delta-doping with Te as co-dopant (for this system, a material used to help in incorporating the dopant); this system has given net acceptor concentrations up to $6 \times 10^{18} \text{ cm}^{-3}$ [35.24] with very low Te concentrations, so that minimal degradation is expected. The second system is that of Li_3N doping, with a report of carrier concentrations close to 10^{19} cm^{-3} . We list recent work on p-type dopants other than N in Table 35.1.

Interestingly, there are two reports that Sb gives quite low activation energies, one being 69 meV [35.25] and the other being 55 meV [35.27] (note that the activation energy for N is 111 meV [35.48]), with the former paper giving a net acceptor concentration of about 10^{16} cm^{-3} ; in this connection it should still be noted that, as mentioned, As and P are generally believed to form DX centers and give deep levels (for a summary [35.10]). Other dopants used were K and Na, with doping carried out via excimer laser annealing; high doping levels were reported, but the excimer procedure would be expected to introduce high defect densities and resultant strong degradation (note that the maximum equilibrium solubility for Na was predicted to be about $5 \times 10^{17} \text{ cm}^{-3}$ by Van de Walle et al. [35.22]). A further approach was that of co-doping, where the term in this case means incorporation of both donors and acceptors; here, experimental tests were reported for Li with I in one case, and with Cl in another, but in both cases net acceptor concentrations were only in the 10^{16} cm^{-3} range. An additional method was to use planes of p-type GaAs (doped with Zn) to inject holes into ZnSe; net acceptor concentrations of 10^{17} cm^{-3} were reported in [35.32], where metalorganic molecular beam epitaxy (MOMBE) was used.

For completeness, we also list in Table 35.3 recent results on n-type doping.

35.1.2 Results on p-Type Material with N as the Primary Dopant

Recent methods for p-type doping with p or $[n_a - n_d]$ exceeding 10^{18} cm^{-3} have been listed in Table 35.2. Note that all of these use N as the primary dopant. As additional comments we note that growth by metalorganic vapor-phase epitaxy (MOVPE) is now relatively standard, and that a quite comprehensive discussion of this method has recently been given [35.49] (although it must be noted that the “hole concentration” of $3 \times 10^{18} \text{ cm}^{-3}$, given in Table 1 of Prete et al. [35.49] from data given in Fujita et al. [35.50], is in fact the N concentration, with Fujita et al. [35.50] giving p as $8 \times 10^{17} \text{ cm}^{-3}$); in view of this extensive recent paper, we do not discuss MOVPE here, but merely give in Table 35.4 some recent references (not in [35.49]). We further note that the use of a δ -ZnSe/ZnTe superlattice (SL) resulted in average Te concentrations of around 9%, which in turn increases the lattice mismatch between the GaAs substrate and the film, since the ZnTe lattice constant is larger than that of ZnSe. This is expected to lead to degradation problems [35.36].

A novel, interesting approach, which has given net acceptor concentrations up to $6 \times 10^{18} \text{ cm}^{-3}$, is that of incorporating both N as a dopant and Te as a co-dopant into the δ -layer(s) with fractional ZnTe coverage, via molecular-beam epitaxy (MBE) [35.24]; as previously mentioned (Sect. 35.1.1), co-dopant here means a material which aids in the incorporation of the dopant, and it is well known that it is easy to obtain p-type ZnTe [35.51, 52]. Electrochemical capacitance–voltage (E–CV) profiling results for various samples are shown in Fig. 35.1 (Fig. 3 of [35.24]); it can be seen that good doping was obtained when three contiguous layers of N and Te were incor-

Table 35.4 Nitrogen-doped ZnSe grown by metalorganic chemical vapor deposition (MOCVD) or MOVPE

$n_a - n_d \text{ (cm}^{-3}\text{)}$	$E_a \text{ (meV)}$	Comments	Reference
6.7×10^{17}	109	ZnSe:N epilayers were grown on ZnSe substrates by low-pressure MOCVD at 830 K and annealed in Zn saturated vapor. The net acceptor concentration is enhanced	[35.45]
1.2×10^{18}		ZnSe:N grown on GaAs. A radio-frequency (RF) plasma nitrogen source was used for doping	[35.35]
		ZnSe:N grown by MOVPE with hydrazines as dopants. The acceptor concentration is limited by the residual impurities in the sources	[35.46]
		ZnSe:N was grown by photo-assisted MOVPE. Post-growth annealing is critical to reducing the hydrogen concentration (by a factor of 10)	[35.47]

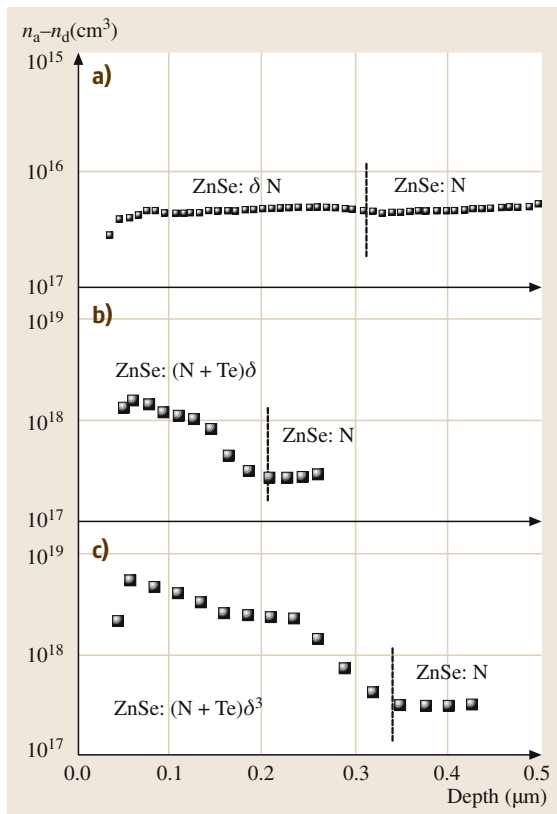


Fig. 35.1 (a) Depth-dependent ($n_a - n_d$) of a conventional δ -doped sample with 5-ML spacer (nominal undoped ZnSe); (b) a (N + Te) δ -doped sample with 4-ML spacer; (c) a (N + Te) δ^3 -doped sample with 7-ML. After [35.24]

porated (δ^3 -doped). A very important aspect of this system, established by subsequent work [35.33], is that the N is preferentially located within ZnTe, which was shown conclusively [35.33] to be present in submonolayer quantities, without formation of a standard SL (considering a standard SL to require full monolayers). This result was established by transmission electron microscopy (TEM), secondary-ion emission spectroscopy (SIMS), and high-resolution X-ray diffraction (HRXRD). SIMS data were taken on a specially prepared sample, in which the spacer regions (undoped ZnSe separating the δ -layers) were thick enough (in view of the SIMS resolution) that the SIMS measurements effectively gave the N and Te concentrations in the delta region. The results for a triple-doped sample are shown in Fig. 35.2, with a Te concentration of about $5 \times 10^{20} \text{ cm}^{-3}$ and an N concentration of about $5 \times 10^{19} \text{ cm}^{-3}$ for a “standard” 5 s Te + N deposition

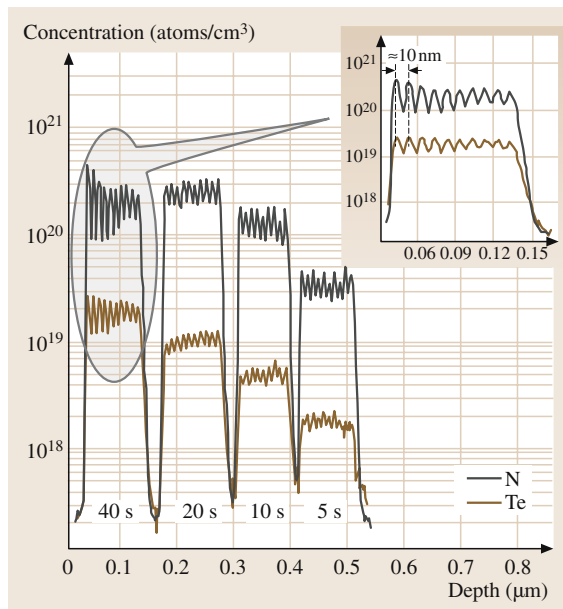


Fig. 35.2 SIMS results on a δ^3 -doped ZnSe:(Te, N) sample. The upper line represents the [Te] concentration, and the lower line represents the [N] concentration

time [35.24]. Thus, both are present at far less than monolayer quantities.

Results from HRXRD are shown in Fig. 35.3, which gives (004) $\theta - 2\theta$ (solid black line) of a triple delta ZnSe:(Te,N) sample; this sample was grown in the [001] direction with a 10 nm ZnSe buffer layer, spacers of 10 monolayer (ML, where we here assume 1 ML, in the [001] growth direction, to be half of the lattice constant), and 200 spacer/ δ -region periods and was grown using a standard Te deposition time of 5 s [35.24]. The strongest peak, at $2\theta \approx 66.01^\circ$, is from the GaAs substrate. In addition, satellite peaks associated with the periodic structure along the growth direction are observed. The result of a simulation using dynamical diffraction theory [35.53–55] is shown by the dashed line. This fit is obtained with a δ -layer and spacer thicknesses of 0.25 ML and 10.4 ML, respectively. These values are in excellent agreement with the nominal growth conditions. The average Te concentrations are $\approx 37\%$ and 2.2% in the δ -layers and the spacers, respectively. The low average Te coverage and its relatively high concentration within the δ^3 -layers indicate that Te is not uniformly distributed within these layers, and, thus, forms ZnTe-rich nano-islands (such nano-islands have been observed by Gu et al. [35.56], optically, in similar samples grown without nitrogen). The relatively

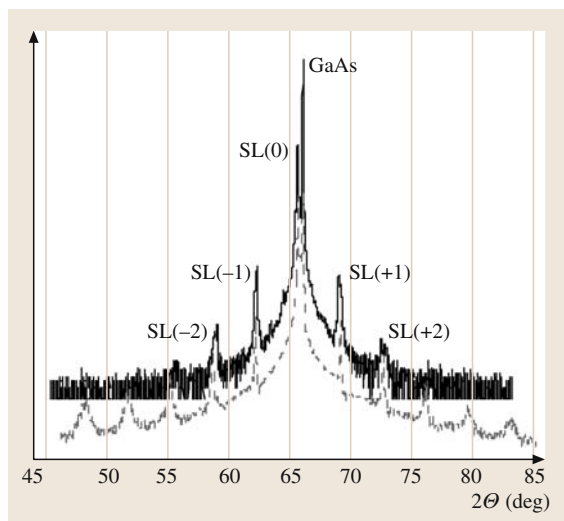


Fig. 35.3 Symmetric $\theta-2\theta$ scan of a δ^3 -doped ZnSe: (Te,N). The solid black line is the experimental result, and the dashed line is the result of simulation. For clarity, the curves are shifted vertically

high Te concentration in the δ -layers is consistent with doping results obtained for ZnSeTe alloys, where it was shown that high acceptor concentrations are observed only for Te concentrations exceeding 15% [35.57]. A further important result for understanding the doping mechanism in the present system was that the photoluminescence (PL) quenched, with increasing temperature, with quite a low activation energy [35.33]; results are shown in Table 35.5, where it can be seen that the activation energies are far lower than for N in ZnSe (111 meV [35.48]) and decrease with increasing Te concentration, down to 38 meV, which is within the range 30–65 meV reported for ZnTe [35.36,58]. Thus, the N is associated primarily with ZnTe, i. e., the N is embedded primarily in Te-rich regions. It can also be noted from the SIMS results (Fig. 35.2) that N and Te are located in the same spatial region, and this is of course totally consistent with the view that N is embedded in Te-rich nano-islands.

We next consider the case of doping by diffusion of Li_3N into MOVPE-grown material [35.34]. The view

has been expressed [35.59] that the resultant good doping was due to the incorporation of Li into Zn sites, and N into Se sites, with both such species being acceptors. In our view, this conjecture is unlikely. Thus, we note that in hard-to-dope wide-band-gap materials, strong compensation is expected [35.60]; since interstitial Li is a donor, it seems very likely that considerable Li is incorporated into the interstitial site after diffusion. Moreover, it is known [35.20] that Li diffuses very quickly. Thus we suggest that, during contact formation, even with minimal heating, a good fraction of the interstitial Li diffuses into precipitates, leaving the material p-type. We note that this view is reinforced by the work of Strassburg et al. [35.61], who show that this method works very well if doping is carried out by ion implantation; such implantation is expected to cause a high density of lattice defects, where defects would be expected to act as nucleation sites for precipitation of interstitial Li.

Last, but not least, no discussion of N doping would be complete without pointing out that it is now realized that N, to a greater or lesser extent (depending on conditions), does self-compensate, i. e. it does introduce donors. The nature of the donors will depend on the Fermi level and on the Zn (Se) and N chemical potentials, as shown in theoretical analyses by Van de Walle et al. [35.22], Kwak et al. [35.62] and Cheong et al. [35.63]. A discussion and comparison of these papers, as well as of the minimum requirements for reliable first-principles calculations, has been given by Neumark [35.10]. Additional work by Faschinger et al. [35.64] and Gundel and Faschinger [35.65] suggested, based on first-principles calculations, a complex between interstitial N (N_i) and a Se vacancy (V_{Se}), but no dependence on the Fermi level or chemical potentials was given.

Moreover, experimentally, Kuskovsky et al. [35.66] have reported a double N interstitial donor at high N doping, and Desgardin et al. [35.67] have reported a $[\text{V}_{\text{Se}}\text{N}_{\text{Se}}]$ complex and a V_{Zn} point defect. Furthermore, it has also been shown that the N_i species (and probably complexes) contribute to degradation [35.11,12,68], but details of this process do not yet appear well understood, and we thus merely mention its existence. However,

Table 35.5 Sample parameters and photoluminescence properties of δ^3 (Te, N)-doped ZnSe

Te concentration (%)	$n_a - n_d$ (cm^{-3})	PL quenching activation energy (meV)
< 3	6.0×10^{18}	38
1.3	4.0×10^{18}	72
< 1	3.0×10^{18}	87

a point we do want to emphasize in this regard is that it might be highly advantageous to be able to use a dopant other than N. We thus note that, with the approach of Lin et al. [35.24], one can envision the use of a dopant

other than N, since the acceptors are not located within ZnSe, but rather in a favorable ZnTe-rich environment. We thus point out that P and As are excellent p-dopants in ZnTe [35.69, 70].

35.2 ZnBeSe

As mentioned in Sect. 35.1, the best p-type dopant developed to date for ZnSe is nitrogen, and such N-doped material suffers from degradation problems. To alleviate this problem, the use of ZnBeSe has been suggested [35.15, 16]. BeSe is *harder* than ZnSe (Fig. 35.4 [35.71]) and, since it is expected that harder materials are less susceptible to defect formation (dislocations etc.), it is expected to be less susceptible to degradation [35.15, 16]. It has been shown that the hardness of ZnBeSe increases with increasing Be content, at least up to 60% Be, as shown in Fig. 35.4 [35.71], where it should be noted that the experimental error at the higher Be concentrations is quite large (moreover, the main interest in ZnBeSe is in the direct-band range, i.e. below 46% [35.72]).

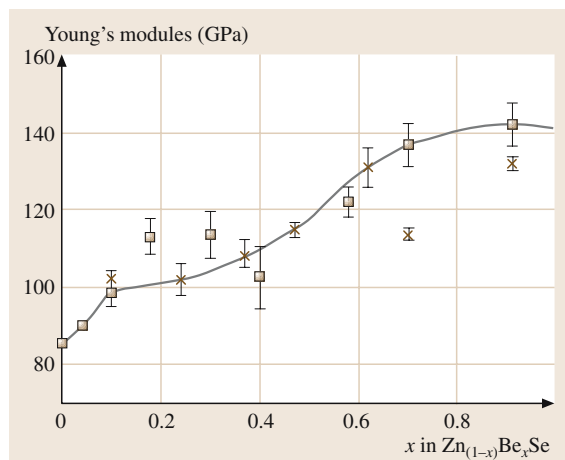


Fig. 35.4 Variation of elastic modulus E as a function of alloy composition x in $\text{ZnBe}_x\text{Se}_{1-x}$. The data points represent the average value. The *squares* – joined by the *full line* – show results obtained under peak loads of 1 mN for alloys grown onto GaAs. *Crosses* are for data obtained under 10 mN for alloys grown onto GaAs and *open circles* show data obtained under 1 mN for alloys grown onto GaP. We note that in general Young's modulus is related to material hardness [35.17]. After [35.71]

It can be pointed out that two additional advantages of ZnBeSe over ZnSe are that one can adjust the lattice constant for better lattice-matching to various materials of interest (GaAs – the most frequently used substrate), and that one can obtain a wider band gap. For instance, $\text{Zn}_{0.028}\text{Be}_{0.972}\text{Se}$ is lattice-matched to GaAs [35.73] and $\text{Zn}_{0.55}\text{Be}_{0.45}\text{Se}$ [35.74] is lattice-matched to Si (assuming a BeSe lattice constant of $\approx 5.138 \text{ \AA}$). The variation of the band gap has been studied in a number of recent papers [35.72, 75]. One result, over the entire concentration range, is shown in Fig. 35.5 [35.72]. It can be seen that the band gap becomes indirect for Be concentrations above 46%; thus, very high Be concentrations are not as interesting, since they cannot be used to give diode lasers.

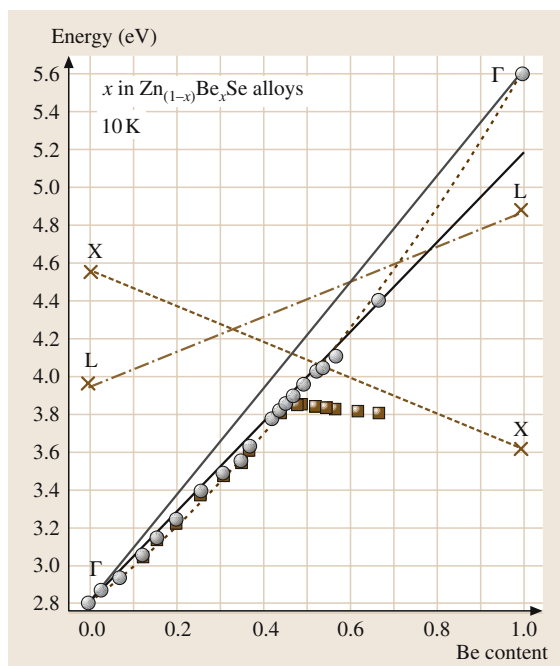


Fig. 35.5 Evolution of the direct band gap (\circ) and of the main PL peak (\blacksquare) as a function of the Be content in $\text{Zn}_{1-x}\text{Be}_x\text{Se}$ alloys. After [35.72]

Table 35.6 XRD, EPD, and C–V results for undoped, N-doped, and (N + Te) δ -doped ZnBeSe epilayers [35.76]

	ZnBeSe	ZnBeSe : N	ZnBeSe : (N + Te) δ	ZnBeSe : (N + Te) δ^3
FWHM (arcs)	23	30	45	51
Be content (%)	3.1	2.6	2.6	2.5
Te content (%)	0	0	0.3	0.5
EPD (cm^{-2})	4×10^4	1×10^5	6×10^5	5×10^5
$n_a - n_d$ (cm^{-3})	–	2×10^{17}	3×10^{17}	1.5×10^{18}

A problem for ZnBeSe, as for ZnSe, is that of p-type doping. The highest bulk net acceptor concentration in p-type ZnBeSe does not exceed $\approx 2 \times 10^{17} \text{ cm}^{-3}$ [35.76]. The best p-type results were again obtained via delta-

doping [35.76], using the same method that Lin et al. [35.24] used for ZnSe. The results from Guo et al. [35.76] are shown in Table 35.6 (Table I from [35.76]).

35.3 ZnO

ZnO is a wide-band-gap (3.36 eV at room temperature) semiconductor with a bulk exciton-binding energy (60 meV), larger than the room-temperature thermal energy, which makes this material very suitable for a variety of applications (see recent reviews by Pearton et al. [35.18], Heo et al. [35.87] and Djurišić et al. [35.88]) in the UV spectral range. However, as for ZnSe and ZnBeSe, one of the major problems for ZnO is p-type doping, and we shall therefore emphasize this aspect.

35.3.1 Doping

ZnO can be grown by a wide range of techniques (some of which are listed in Table 35.7). As-grown ZnO is usually n-type, and heavily n-type ZnO is easily obtained by using group III elements. It is assumed that nominally undoped ZnO is n-type due to shallow native defects such as interstitial zinc (Zn_i) [35.89, 90] or, alternatively, due to the presence of hydrogen [35.91]. Experimentally, hydrogen in ZnO has been observed

Table 35.7 p-type doping of ZnO

Dopant	Growth method	Resistivity (Ωcm)	Dopant concentration (cm^{-3})	Carrier concentration (cm^{-3})	Mobility (cm^2/Vs)	Reference
As	Evaporation followed by sputtering	0.4	Mid 10^{19}	4×10^{18}	4	[35.77]
As	Hybrid beam deposition	2	3×10^{18}	4×10^{17}	35	[35.78]
P	RF sputtering followed by RTA	0.59–4.4		1.0×10^{17} – 1.7×10^{19}	0.53–3.51	[35.79]
N	Ultrasonic spray pyrolysis	≈ 0.03		8.59×10^{18}	24.1	[35.80]
N	Thermal oxidization of Zn_3N_2 thin films		Up to 10^{21}	4.16×10^{17}		[35.81]
N	Implantation	10.11–15.3		Up to 7.3×10^{17}	2.51–6.02	[35.82]
N	MOCVD	3.02		1.97×10^{18}	1	[35.83]
N	CVD	17.3	Up to 3×10^{20}	1.06×10^{18}	0.34	[35.84]
N + Al	Direct-current (DC) reactive magnetron sputtering	57.3		2.25×10^{17}	0.43	[35.85]
N + In	Ultrasonic spray pyrolysis	0.017		2.44×10^{18}	155	[35.86]

via electron paramagnetic resonance (EPR), electron nuclear double resonance (ENDOR), optical, and IR absorption measurements [35.92–95]. The activation energy of the hydrogen donor is 35–46 meV [35.92, 96]. We note that sometimes oxygen vacancies (V_O) are cited as shallow donors [35.97]; however, *Zhang et al.* [35.90] estimated this species to be a relatively deep level. Also, *Vanheusden et al.* [35.98, 99] suggested that charged oxygen vacancies are responsible for the deep green luminescence in ZnO (see below).

Obtaining good p-type ZnO has however proven difficult. There is a good discussion and summary of growth methods as well as achieved resistivities in p-type ZnO up to 2003 in *Look and Claffin* [35.9] and *Look et al.* [35.102]. The latter publication also discusses background impurities in ZnO. We therefore present only some later results and give a short discussion of models proposed for p-type doping of ZnO.

Group V acceptors, based on theoretical arguments, are expected to form very deep substitutional acceptors; for instance, *Park et al.* [35.103] have calculated that the ionization energies of N, P, and As are 0.40 eV, 0.93 eV, and 1.15 eV, respectively. So successes (Table 35.7) in obtaining p-type ZnO with N, P, and As are surprising. Also group I (Li, Na, and K) impurities [35.103] have, in general, lower ionization energies, but these impurities are amphoteric and thus self-compensate. Experimentally, interstitial Li and Na donors were observed by *Orlinskii et al.* [35.104], and recent attempt to use Li_3N to dope ZnO to be p-type produced n-type conductivity instead [35.105].

To achieve p-type doping, *Wang and Zunger* [35.106] have proposed a cluster co-doping method using Ga or Al as co-dopants along with group V dopants; experimentally, p-type ZnO has been obtained using co-doping with Al and In (Table 35.7); N–Ga co-doping has been attempted [35.107] but no p-type conductivity has been observed via the Hall effect.

Recently, to explain p-type ZnO obtained via group V doping, *Limpijumnong et al.* [35.108] proposed, using first-principles calculations, that group V elements give shallow acceptors by forming complexes with native defects. Specifically, these authors proposed that $ZnO:As$ and $ZnO:Sb$ are p-type due to $As_{Zn}-2V_{Zn}$ and $Sb_{Zn}-2V_{Zn}$ complexes, which behave as shallow acceptors. These complexes have low formation energies (1.59 eV and 2.00 eV, respectively) as well as low ionization energies (0.15 eV and 0.16 eV, respectively). Experimentally, the activation energy for $ZnO:As$ was reported to be between 0.12 [35.78] and 0.18 [35.109]. As for nitrogen, the most often used p-type dopant,

Look et al. [35.77] reported that the ionization energy was as low as 0.090 eV for heavily doped material (see also [35.110] and references therein).

Regarding doping using phosphorous, we note that *Kim et al.* [35.79] obtained p-type ZnO only after annealing at high temperatures using rapid thermal annealing (RTA), while as-grown material was n-type. The authors suggested that the annealing removes the compensating donors; however, we suggest that the formation of shallow acceptor complexes cannot be ruled out, especially in view of enhanced n-type behavior with increased P concentration [35.87].

Finally, we note that *Lee and Chang* [35.111] have proposed, theoretically, ways to use [group I – Hydrogen] complexes for p-type doping. These authors have found that an intentional co-doping with H impurities suppresses the formation of compensating interstitials and greatly enhances the solubility of Li and Na acceptors. This type of effect, in general, was clearly predicted by *Neumark* [35.112]. H atoms can be easily removed from ZnO by post-growth annealing at relatively low temperatures. Apparently, this method is similar to that used to obtain p-type GaN. These authors [35.111] also found, as did *Park et al.* [35.103], that Li and Na have lower ionization energies than substitutional group V dopants such as nitrogen.

35.3.2 Optical Properties

Finally, we shall briefly discuss some optical properties of ZnO. Low-temperature PL of undoped ZnO is dominated by near-band, edge emission, with up to 20 lines observed within the spectral range 3.34–3.38 eV [35.96]. Detailed studies of bound excitons (BX) and donor–acceptor pair luminescence have recently been published by *Meyer et al.* [35.96], so here

Table 35.8 Low-temperature bound-excitonic position and assignments

BX line energy (eV)	Assignment	Donor binding energy (meV)	References
≈ 3.3567	In	63.2	[35.96, 100]
≈ 3.3598	Ga	54.6	[35.96, 101]
≈ 3.3628	H	37 35 46.1	[35.92] [35.95] [35.96]
≈ 3.3608	Al	54.8 51.55	[35.96] [35.100]

we merely summarize some of their results and compare these with other reports.

An important conclusion was that all bound-excitonic lines are due to neutral donor-bound excitons, rather than, as previously suggested, to acceptor-bound excitons [35.113–115]. For instance, the ≈ 3.357 eV and ≈ 3.3608 lines are attributed to In and Al donors, respectively [35.113–115]. Similar conclusions were obtained by *Morhain* et al. [35.100] using PL and selective PL (SPL) on MBE-grown ZnO. It must be noted that such assignments were previously proposed by *Block* et al. [35.116] and *Gonzalez* et al. [35.117]. In Table 35.8 we summarize the BX lines that have been identified with a specific impurity or center. A line at 3.3631 eV, which is slightly above the 3.3628 eV line, was also assigned to hydrogen by *Look* et al. [35.95] who performed Hall-effect, PL, and EPR measurements on a series of ZnO samples annealed in air at various temperatures. The dominant donor had an activation energy of ≈ 37 meV, but disappeared after high-temperature annealing, and was replaced by a 67 meV donor [35.95]. The line at ≈ 3.3631 eV has been assigned to the 37 meV donor; the authors suggested, following *Hofmann* et al. [35.92], that this donor is hydrogen. This assignment has been confirmed by *Meyer* et al. [35.96] by SPL and by *Morhain* et al. [35.100] by magneto-optics, with *Meyer* et al. [35.96] reporting an H ionization energy of ≈ 46.1 meV.

Another important feature of bulk ZnO is a visible luminescence, often referred to as the green band. It is

usually observed at (2.38 ± 0.04) eV [35.118–121]. The origin of this band, however, remains controversial: transitions associated with O_{Zn} antisites [35.119], oxygen vacancies [35.97–99, 122], zinc interstitials [35.123], ZnO antisites [35.124], donor–acceptor pairs [35.125], and Cu^{2+} ions [35.126] have all been suggested. It must be noted that the origin of the green luminescence could be different in ZnO prepared via different methods, since various defects and/or impurities can contribute to the emission [35.127]. We note that oxygen vacancies are the species most often suggested as the defect associated with the green luminescence. Oxygen vacancies can have three states – neutral, singly and doubly positively charged. The transition thus depends on the type of free carrier that is participating in recombination. *Vanheusden* et al. [35.98, 99] suggested that holes participate in this recombination while *Djurišić* et al. [35.88] (also references therein) suggested the involvement of electrons. We have recently shown that the green luminescence most likely involves holes rather than electrons [35.128], via studies of quantum ZnO wires. A similar conclusion was also recently obtained by *Kang* et al. [35.129] who investigated PLD-grown ZnO. Their conclusion was that the most likely channel for the green PL is through a deep donor (attributed to oxygen vacancies) and the holes in the valence bands. Lastly, a further suggestion by *Harada* and *Hashimoto* [35.130] is that surface states associated with oxygen vacancies could play a significant role in the emission within the visible spectral region in bulk ZnO.

References

- 35.1 L. Svob, C. Thiandourme, A. Lusson, M. Bouanani, Y. Marfaing, O. Gorochoy: Appl. Phys. Lett. **76**, 1695 (2000)
- 35.2 S. Kishimoto, T. Hasegawa, H. Kinto, O. Matsumoto, S. Iida: J. Cryst. Growth **214/215**, 556 (2000)
- 35.3 S. Kishimoto, A. Kato, A. Naito, Y. Yakamoto, S. Iida: Phys. Status Solidi B **229**, 391 (2002)
- 35.4 Y. Abiko, N. Nakayama, K. Akimoto, T. Yao: Phys. Status Solidi B **229**, 339 (2001)
- 35.5 S. Nakamura, J. Yamaguchi, S. Takagimoto, Y. Yamada, T. Taguchi: J. Cryst. Growth **237/239**, 1570 (2002)
- 35.6 S. Kohiki, T. Suzuka, M. Oku, T. Yamamoto, S. Kishimoto, S. Iida: J. Appl. Phys. **91**, 760 (2002)
- 35.7 K. Ichino, Y. Matsuki, S. T. Lee, T. Nishikawa, M. Kitagawa, H. Kobayashi: Phys. Status Solidi C **1**, 710 (2004)
- 35.8 S. J. Pearton, D. P. Norton, K. Lp, Y. W. Heo, T. Steiner: J. Vac. Sci. Technol. B **22**, 932 (2004)
- 35.9 D. C. Look, B. Claflin: Phys. Status Solidi B **241**, 624 (2004)
- 35.10 G. F. Neumark: Mater. Lett. **30**, 131 (1997)
- 35.11 D. Albert, J. Nürnberger, V. Hock, M. Ehinger, W. Faschinger, G. Landwehr: Appl. Phys. Lett. **74**, 1957 (1999)
- 35.12 V. N. Jmerik, S. V. Sorokin, T. V. Shubina, N. M. Shmidt, I. V. Sedova, D. L. Fedorov, S. V. Ivanov, P. S. Kop'ev: J. Cryst. Growth **214/215**, 502 (2000)
- 35.13 H. Ebe, B.-P. Zhang, F. Sakurai, Y. Segawa, K. Suto, J. Nishizawa: Phys. Status Solidi B **229**, 377 (2002)
- 35.14 K. Katayama, T. Nakamura: J. Appl. Phys. **95**, 3576 (2004)
- 35.15 A. Wagg, F. Fischer, H.-J. Lugauer, Th. Litz, T. Gerhard, J. Nürnberger, U. Lünz, U. Zehnder, W. Ossau, G. Landwehr, B. Roos, H. Richter: Mater. Sci. Eng. B **43**, 65 (1997)
- 35.16 C. Verie: J. Cryst. Growth **184/185**, 1061 (1998)

- 35.17 F. C. Peiris, U. Bindley, J. K. Furdyna, H. Kim, A. K. Raudas, M. Grimsditch: *Appl. Phys. Lett.* **79**, 473 (2001)
- 35.18 S. J. Pearton, C. R. Abernathy, M. E. Overberg, G. T. Thaler, D. P. Northon, N. Theodoropoulou, A. F. Hebard, Y. D. Park, F. Ren, J. Kim, L. A. Boatner: *J. Appl. Phys.* **93**, 1 (2003)
- 35.19 Y. W. Heo, D. P. Norton, L. C. Tien, Y. Kwon, B. S. Kang, F. Ren, S. J. Pearton, J. R. LaRoche: *Mater. Sci. Eng. R* **47**, 1 (2004)
- 35.20 M. A. Haase, H. Cheng, J. M. DePuydt, J. E. Potts: *J. Appl. Phys.* **67**, 448 (1990)
- 35.21 G. F. Neumark, S. P. Herko: *J. Cryst. Growth* **59**, 189 (1982)
- 35.22 C. G. Van de Walle, D. B. Laks, G. F. Neumark, S. T. Pantelides: *Phys. Rev. B* **47**, 9425 (1993)
- 35.23 G. F. Neumark, S. P. Herko, T. F. McGee III, B. J. Fitzpatrick: *Phys. Rev. Lett.* **53**, 604 (1984)
- 35.24 W. Lin, S. P. Guo, M. C. Tamargo, I. Kuskovsky, C. Tian, G. F. Neumark: *Appl. Phys. Lett.* **76**, 2205 (2000)
- 35.25 M. Takemura, H. Goto, T. Ido: *Jpn. J. Appl. Phys.* **36**, L540 (1997)
- 35.26 H. Goto, T. Ido, A. Takatasuka: *J. Cryst. Growth* **214/215**, 529 (2000)
- 35.27 H. Kalisch, H. Hamadeh, R. Rüländ, J. Berntgen, A. Krysa, M. Hluken: *J. Cryst. Growth* **214/215**, 1163 (2000)
- 35.28 M. Prokesch, K. Irmscher, U. Rinas, H. Makino, T. Yao: *J. Cryst. Growth* **242**, 155 (2002)
- 35.29 Y. Hatanaka, M. Niraula, A. Nakamura, T. Aoki: *Appl. Surf. Sci.* **175/176**, 462 (2001)
- 35.30 I. Suemune, H. Ohsawa, T. Tawara, H. Machida, N. Shimoyama: *J. Cryst. Growth* **214/215**, 562 (2000)
- 35.31 M. Yoneta, H. Uechi, K. Nanami, M. Ohishi, H. Saito, K. Yoshino, K. Ohmori: *Physica B* **302**, 166 (2001)
- 35.32 J. Hirose, I. Suemune, A. Ueta, H. Machida, N. Shimoyama: *J. Cryst. Growth* **214/215**, 524 (2000)
- 35.33 I. L. Kuskovsky, Y. Gu, Y. Gong, H. F. Yan, J. Lau, G. F. Neumark, O. Maksimov, X. Zhou, M. C. Tamargo, V. Volkov, Y. Zhu, L. Wang: *Phys. Rev. Lett. B* **73**, 195306 (2006)
- 35.34 O. Schulz, M. Strassburg, T. Rissom, U. W. Pohl, D. Bimberg, M. Klude, D. Hommel: *Appl. Phys. Lett.* **81**, 4916 (2002)
- 35.35 E. D. Sim, Y. S. Joh, J. H. Song, H. L. Park, S. H. Lee, K. Jeong, S. K. Chang: *Phys. Status Solidi B* **229**, 213 (2002)
- 35.36 H. D. Jung, C. D. Song, S. Q. Wang, K. Arai, Y. H. Wu, Z. Zhu, T. Yao, H. Katayama-Yoshida: *Appl. Phys. Lett.* **70**, 1143 (1997)
- 35.37 S. Z. Wang, S. F. Yoon, L. He, X. C. Shen: *J. Appl. Phys.* **90**, 2314 (2001)
- 35.38 M. Yoneta, K. Nanami, H. Uechi, M. Ohishi, H. Saito, K. Yoshino: *J. Cryst. Growth* **237/239**, 1545 (2002)
- 35.39 Y. Gu, I. L. Kuskovsky, G. F. Neumark, X. Zhou, O. Maksimov, S. P. Guo, M. C. Tamargo: *J. Lumin.* **104**, 77 (2003)
- 35.40 D. C. Oh, J. S. Song, J. H. Chang, T. Takai, T. Handa, M. W. Cho, T. Yao: *Mater. Sci. Semicond. Process.* **6**, 567 (2003)
- 35.41 D. C. Oh, J. H. Chang, T. Takai, J. S. Song, K. Godo, Y. K. Park, K. Shindo, T. Yao: *J. Cryst. Growth* **251**, 607 (2003)
- 35.42 H. Kato, H. Udono, I. Kikuma: *J. Cryst. Growth* **229**, 79 (2001)
- 35.43 M. Yoneta, T. Kubo, H. Kato, K. Yoshino, M. Ohishi, H. Saito, K. Ohmori: *Phys. Status Solidi B* **229**, 291 (2002)
- 35.44 K. Lott, O. Volobujeva, A. Öpik, T. Nirk, L. Türn, M. Noges: *Phys. Status Solidi C* **0**, 618 (2003)
- 35.45 J. F. Wang, D. Masugata, C. B. Oh, A. Omino, S. Seto, M. Isshikim: *Phys. Status Solidi A* **193**, 251 (2002)
- 35.46 U. W. Pohl, J. Gottfriedsen, H. Schumann: *J. Cryst. Growth* **209**, 683 (2000)
- 35.47 M. U. Ahmed, S. J. C. Irvine: *J. Electron. Mater.* **29**, 169 (2000)
- 35.48 P. J. Dean, W. Stutius, G. F. Neumark, B. J. Fitzpatrick, R. N. Bhargava: *Phys. Rev. B* **27**, 2419 (1983)
- 35.49 P. Prete, N. Lovergine: *Prog. Cryst. Growth Char. Mater.* **44**, 1 (2002)
- 35.50 Y. Fujita, T. Terada, T. Suzuki: *Jpn. J. Appl. Phys.* **34**, L1034 (1995)
- 35.51 C. M. Rouleau, D. H. Lowndes, G. W. McCamy, J. D. Budai, D. B. Poker, D. B. Geohegan, A. A. Puzretsky, S. Zhu: *Appl. Phys. Lett.* **67**, 2545 (1995)
- 35.52 T. Baron, K. Saminadayar, N. Magnea: *J. Appl. Phys.* **83**, 1354 (1998)
- 35.53 S. Takagi: *Acta Crystallogr.* **15**, 1311 (1962)
- 35.54 D. Taupin: *Bull. Soc. Franc. Miner. Crystallogr.* **88**, 469 (1964)
- 35.55 M. A. G. Halliwell, M. H. Lyons, M. J. Hill: *J. Cryst. Growth* **68**, 523 (1984)
- 35.56 Y. Gu, I. L. Kuskovsky, M. van der Voort, G. F. Neumark, X. Zhou, M. C. Tamargo: *Phys. Rev. B* **71**, 045340 (2005)
- 35.57 W. Lin, B. S. Yang, S. P. Guo, A. Elmoumni, F. Fernandez, M. C. Tamargo: *Appl. Phys. Lett.* **75**, 2608 (1999)
- 35.58 N. J. Duddles, K. A. Dhese, P. Devine, D. E. Ashenford, C. G. Scott, J. E. Nicholls, J. E. Lunn: *J. Appl. Phys.* **76**, 5214 (1994)
- 35.59 S. W. Lim, T. Honda, F. Koyama, K. Iga, K. Inoue, K. Yanashima, H. Munekata, H. Kukimoto: *Appl. Phys. Lett.* **65**, 2437 (1994)
- 35.60 G. F. Neumark, R. M. Park, J. M. Depuydt: *Phys. Today* **47** (6), 26 (1994)
- 35.61 M. Strassburg, O. Schulz, U. W. Pohl, D. Bimberg, S. Itoh, K. Nakano, A. Ishibashi, M. Klude, D. Hommel: *IEEE J. Sel. Top. Quant. Electron.* **7**, 371 (2001)
- 35.62 K. W. Kwak, R. D. King-Smith, D. Vanderbilt: *Physica B* **185**, 154 (1993)
- 35.63 B.-H. Cheong, C. H. Park, K. J. Chang: *Phys. Rev. B* **51**, 10610 (1995)

- 35.64 W. Faschinger, S. Gundel, J. Nürnberger, D. Albert: *Proc. Conf. Optoelectronic and Microelectronic Materials and Devices* (IEEE, Piscataway 2000) p. 41
- 35.65 S. Gundel, W. Faschinger: *Phys. Rev. B* **65**, 035208 (2001)
- 35.66 I. L. Kuskovsky, G. F. Neumark, J. G. Tischler, B. A. Weinstein: *Phys. Rev. B* **63**, 161201 (2001)
- 35.67 P. Desgardin, J. Oila, K. Sarnen, P. Hautojärvi, E. Tournié, J.-P. Faurie, C. Corbel: *Phys. Rev. B* **62**, 15711 (2000)
- 35.68 S. Tomiya, S. Kijima, H. Okuyama, H. Tsukamoto, T. Hino, S. Taniguchi, H. Noguchi, E. Kato, A. Ishibashi: *J. Appl. Phys.* **86**, 3616 (1999)
- 35.69 F. El. Akkad: *Semicond. Sci. Technol.* **2**, 629 (1987)
- 35.70 A. Kamata, H. Yoshida: *Jpn. J. Appl. Phys. (Pt. 2)* **135**, L87 (1996)
- 35.71 S. E. Grillo, M. Ducarrori, M. Nadal, E. Tournié, J.-P. Faurie: *J. Appl. Phys. D: Appl. Phys.* **35**, 3015 (2002)
- 35.72 C. Chauvet, E. Tournié, J.-P. Faurie: *Phys. Rev. B* **61**, 5332 (2000)
- 35.73 V. Bousquet, E. Tournié, M. Läugt, P. Venéguès, J.-P. Faurie: *Appl. Phys. Lett.* **70**, 3564 (1997)
- 35.74 J. P. Faurie, V. Bousquet, P. Brunet, E. Tournié: *J. Cryst. Growth* **184/185**, 11 (1998)
- 35.75 M. Malinski, L. Bychto, S. Legowski, J. Szatkowski, J. Zakrzewski: *Microelectron. J.* **32**, 903 (2001)
- 35.76 S. P. Guo, W. Lin, X. Zhou, M. C. Tamargo, C. Tian, I. L. Kuskovsky, G. F. Neumark: *J. Appl. Phys.* **90**, 1725 (2001)
- 35.77 D. C. Look, G. M. Renlund, R. H. Burgener II, J. R. Sizelove: *Appl. Phys. Lett.* **85**, 5268 (2004)
- 35.78 Y. R. Ryu, T. S. Lee, H. W. White: *Appl. Phys. Lett.* **83**, 87 (2003)
- 35.79 K.-K. Kim, H. S. Kim, D.-K. Hwang, J.-H. Lim, S.-J. Park: *Appl. Phys. Lett.* **83**, 63 (2003)
- 35.80 J. M. Bian, X. M. Li, C. Y. Zhang, W. D. Yu, X. D. Gao: *Appl. Phys. Lett.* **85**, 4070 (2004)
- 35.81 B. S. Li, Y. C. Liu, Z. Z. Zhi, D. Z. Shen, Y. M. Lu, J. Y. Zhang, X. W. Fan, R. X. Mu, D. O. Henderson: *J. Mater. Res.* **18**, 8 (2003)
- 35.82 C. C. Lin, S. Y. Shen, S. Y. Cheng, H. Y. Li: *Appl. Phys. Lett.* **84**, 5040 (2004)
- 35.83 W. Xu, Z. Ye, T. Zhou, B. Zhao, L. Zhu, J. Huang: *J. Cryst. Growth* **265**, 133 (2004)
- 35.84 X. Li, Y. Yan, T. A. Gessert, C. L. Perkins, D. Young, C. DeHart, M. Young, T. J. Coutts: *J. Vac. Sci. Technol. A* **21**, 1342 (2003)
- 35.85 J. G. Lu, Z. Z. Ye, F. Zhuge, Y. J. Zeng, B. H. Zhao, L. P. Zhu: *Appl. Phys. Lett.* **85**, 3134 (2004)
- 35.86 J. M. Bian, X. M. Li, X. D. Gao, W. D. Yu, L. D. Chen: *Appl. Phys. Lett.* **84**, 541 (2004)
- 35.87 Y. W. Heo, K. Ip, S. J. Park, S. J. Peaton, D. P. Norton: *Appl. Phys. A* **78**, 53 (2004)
- 35.88 A. B. Djurišić, Y. Chan, E. H. Li: *Mater. Sci. Eng. R* **38**, 237 (2002)
- 35.89 D. C. Look, J. W. Hemsky, J. R. Sizelove: *Phys. Rev. Lett.* **82**, 2552 (1999)
- 35.90 S. B. Zhang, S.-H. Wei, A. Zunger: *Phys. Rev. B* **63**, 075205 (2001)
- 35.91 C. G. Van de Walle: *Phys. Rev. Lett.* **85**, 1012 (2000)
- 35.92 D. M. Hofmann, A. Hofstaetter, F. Leiter, H. Zhou, F. Henecker, B. K. Meyer, S. B. Orlinskii, J. Schmidt, P. G. Baranov: *Phys. Rev. Lett.* **88**, 045504 (2002)
- 35.93 E. V. Lavrov, J. Weber, F. Börnert, C. G. Van de Walle, R. Helbig: *Phys. Rev. B* **66**, 165205 (2001)
- 35.94 M. D. McCluskey, S. J. Jokela, K. K. Zhuravlev, P. J. Simpson, K. G. Lynn: *Appl. Phys. Lett.* **81**, 3807 (2002)
- 35.95 D. C. Look, R. L. Jones, J. R. Sizelove, N. Y. Garces, N. C. Giles, L. E. Halliburton: *Phys. Status Solidi A* **195**, 171 (2003)
- 35.96 B. K. Meyer, H. Alves, D. M. Hofmann, W. Kriegseis, D. Forster, F. Bertram, J. Christen, A. Hoffmann, M. Straßburg, M. Dworzak, U. Haboeck, A. V. Rodina: *Phys. Status Solidi B* **241**, 231 (2004)
- 35.97 F. A. Kroger, H. J. Vink: *J. Chem. Phys.* **22**, 250 (1954)
- 35.98 K. Vanheusden, C. H. Seager, W. L. Warren, D. R. Tallent, J. A. Voigt: *Appl. Phys. Lett.* **68**, 403 (1996)
- 35.99 K. Vanheusden, W. L. Warren, C. H. Seager, D. R. Tallent, J. A. Voigt: *J. Appl. Phys.* **79**, 7983 (1996)
- 35.100 C. Morhain, M. Teisseire-Doninelli, S. Vézian, C. Deparis, P. Lorenzini, F. Raymond, J. Guion, G. Neu: *Phys. Status Solidi B* **241**, 631 (2004)
- 35.101 H. J. Ko, Y. F. Chen, S. K. Hong, H. Wenisch, T. Yao, D. C. Look: *Appl. Phys. Lett.* **77**, 3761 (2000)
- 35.102 D. C. Look, B. Clafin, Ya. I. Alivov, S. J. Park: *Phys. Stat. Sol. A* **201**, 2203 (2004)
- 35.103 C. H. Park, S. B. Zhang, S.-H. Wei: *Phys. Rev. B* **66**, 073202 (2002)
- 35.104 S. Orlinskii, J. Schmdit, P. G. Baranov, D. M. Hofmann, C. de M. Donegá, A. Meijerink: *Phys. Rev. Lett.* **92**, 047603 (2004)
- 35.105 H.-J. Ko, Y. Chen, S.-K. Hong, T. Yao: *J. Cryst. Growth* **251**, 628 (2003)
- 35.106 L. G. Wang, A. Zunger: *Phys. Rev. Lett.* **90**, 256401 (2003)
- 35.107 M. Sumiya, A. Tsukazaki, S. Fuke, A. Ohtomo, H. Koinuma, M. Kawasaki: *Appl. Surf. Sci.* **223**, 206 (2004)
- 35.108 S. Limpijumnon, S. B. Zhang, S.-H. Wei, C. H. Park: *Phys. Rev. Lett.* **92**, 155504 (2004)
- 35.109 C. Morhain, M. Teisseire, S. Vézian, F. Vigué, F. Raymond, P. Lorenzini, J. Guion, G. Neu, J.-P. Faurie: *Phys. Status Solidi B* **229**, 881 (2002)
- 35.110 D. C. Look, B. Clafin, Ya. I. Alivov, S. J. Park: *Phys. Status Solidi A* **201**, 2203 (2004)
- 35.111 E.-C. Lee, K. J. Chang: *Phys. Rev. B* **70**, 115210 (2004)
- 35.112 G. F. Neumark: *Phys. Rev. Lett.* **62**, 1800 (1989)
- 35.113 G. Blattner, C. Klingshirn, R. Helbig, R. Meinel: *Phys. Status Solidi B* **107**, 105 (1981)
- 35.114 C. Klingshirn, W. Maier, G. Blatter, P. J. Dean, G. Klobbe: *J. Cryst. Growth* **59**, 352 (1982)

- 35.115 J. Gutowski, N. Presser, I. Broser: Phys. Rev. B **38**, 9746 (1988)
- 35.116 D. Block, A. Hervé, R. T. Cox: Phys. Rev. B **25**, 6049 (1982)
- 35.117 C. Gonzalez, D. Block, R. T. Cox, A. Hervé: J. Cryst. Growth **59**, 357 (1982)
- 35.118 X. Liu, X. Wu, H. Cao, R. P. H. Chang: J. Appl. Phys. **95**, 3141 (2004)
- 35.119 B. Lin, Z. Fu, Y. Jia: Appl. Phys. Lett. **79**, 943 (2001)
- 35.120 D. Banejee, J. Y. Lao, D. Z. Wang, J. Y. Huang, Z. F. Ren, D. Steeves, B. Kimball, M. Sennett: Appl. Phys. Lett. **83**, 2061 (2003)
- 35.121 T.-B. Hur, G. S. Jeon, Y.-H. Hwang, H.-K. Kim: J. Appl. Phys. **94**, 5787 (2003)
- 35.122 P. H. Hasai: Phys. Rev. **130**, 989 (1963)
- 35.123 M. Liu, A. H. Kitai, P. Mascher: J. Lumin. **54**, 35 (1992)
- 35.124 D. C. Reynolds, S. C. Look, B. Jogai, H. Morkoc: Solid State Commun. **101**, 643 (1997)
- 35.125 D. C. Reynolds, S. C. Look, B. Jogai: J. Appl. Phys. **89**, 6189 (2001)
- 35.126 N. Y. Garces, L. Wang, L. Bai, N. C. Giles, I. E. Haliburton, G. Cantwell: Appl. Phys. Lett. **81**, 622 (2002)
- 35.127 D. Li, H. Leung, A. B. Djurišić, Z. T. Liu, M. H. Xie, S. L. Shi, S. J. Xu, W. K. Chan: Appl. Phys. Lett. **85**, 1601 (2004)
- 35.128 Y. Gu, I. L. Kuskovsky, M. Yin, S. O'Brien, G. F. Neumark: Appl. Phys. Lett. **85**, 3833 (2004)
- 35.129 H. S. Kang, J. S. Kang, J. W. Kim, S. Y. Lee: Phys. Status Solidi C **1**, 2550 (2004)
- 35.130 Y. Harada, S. Hashimoto: Phys. Rev. B **68**, 045421 (2003)

36. II–VI Narrow-Bandgap Semiconductors for Optoelectronics

The field of narrow-gap II–VI materials is dominated by the compound semiconductor mercury cadmium telluride, ($\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ or MCT), which supports a large industry in infrared detectors, cameras and infrared systems. It is probably true to say that HgCdTe is the third most studied semiconductor after silicon and gallium arsenide. $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is the material most widely used in high-performance infrared detectors at present. By changing the composition x the spectral response of the detector can be made to cover the range from $1\text{ }\mu\text{m}$ to beyond $17\text{ }\mu\text{m}$. The advantages of this system arise from a number of features, notably: close lattice matching, high optical absorption coefficient, low carrier generation rate, high electron mobility and readily available doping techniques. These advantages mean that very sensitive infrared detectors can be produced at relatively high operating temperatures. $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ multilayers can be readily grown in vapor-phase epitaxial processes. This provides the device engineer with complex doping and composition profiles that can be used to further enhance the electro-optic performance, leading to low-cost, large-area detectors in the future. The main purpose of this chapter is to describe the applications, device physics and technology of II–VI narrow-bandgap devices, focusing on HgCdTe but also including $\text{Hg}_{1-x}\text{Mn}_x\text{Te}$ and $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$. It concludes with a review of the research and development programs into third-generation infrared detector technology (so-called GEN III detectors) being performed in centers around the world.

36.0.1	Historical Perspective and Early Detectors	856
36.0.2	Introduction to HgCdTe	857
36.0.3	Introduction to Device Types.....	857

36.1	Applications and Sensor Design	858
36.2	Photoconductive Detectors in HgCdTe and Related Alloys	860
36.2.1	Introduction to the Technology of Photoconductor Arrays	860
36.2.2	Theoretical Fundamentals for Long-Wavelength Arrays	861
36.2.3	Special Case of Medium-Wavelength Arrays....	863
36.2.4	Nonequilibrium Effects in Photoconductors	863
36.3	SPRITE Detectors	864
36.4	Photoconductive Detectors in Closely Related Alloys	866
36.5	Conclusions on Photoconductive HgCdTe Detectors ...	867
36.6	Photovoltaic Devices in HgCdTe	867
36.6.1	Ideal Photovoltaic Devices.....	868
36.6.2	Nonideal Behavior in HgCdTe Diodes.....	869
36.6.3	Theoretical Foundations of HgCdTe Array Technology	870
36.6.4	Manufacturing Technology for HgCdTe Arrays	873
36.6.5	HgCdTe 2–D Arrays for the 3–5 μm (MW) Band	878
36.6.6	HgCdTe 2–D Arrays for the 8–12 μm (LW) Band	879
36.6.7	HgCdTe 2–D Arrays for the 1–3 μm (SW) Band.....	879
36.6.8	Towards “GEN III Detectors”	880
36.6.9	Conclusions and Future Trends for Photovoltaic HgCdTe Arrays	882
36.7	Emission Devices in II–VI Semiconductors	882
36.8	Potential for Reduced-Dimensionality HgTe–CdTe	883
	References	883

The main commercial application for narrow-bandgap semiconductors is in infrared radiation detection. There are very few elemental or compound semiconductors

with the correct energy gap to sense photons within the infrared spectrum, particularly at longer wavelengths. It is also beneficial to accurately match the spectral sen-

sitivity to certain atmospheric windows, so the ability to tailor the wavelength is very important. By using an alloy of two different compounds with widely separate energy gaps it is possible to synthesize crystals with an intermediate energy gap. II–VI compounds have crystal properties that make them very suitable for mixing and they have a range of bandgaps from near zero for the semimetal Hg compounds to more than 1 eV.

However, amongst the wider bandgap II–VI semiconductors, only Cd, Zn, Mn and Mg have been shown to open up the bandgap of the Hg-based semimetals HgTe and HgSe. By far the most developed alloy system is $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (HgCdTe), which is a semiconductor formed from the semimetal HgTe and the wide-bandgap semiconductor CdTe. By adjusting the alloy composition “ x ”, the properties (including the bandgap) can be varied smoothly between HgTe and CdTe.

There has been some interest in developing alternative ternary alloys to replace HgCdTe, since from theoretical considerations the already weak Hg–Te lattice bond is further destabilized by alloying with CdTe and there is potential for obtaining materials with increased hardness and detectors with better temperature stability. Most other compound combinations have crystal growth problems or doping limitations that make them unsuitable for device fabrication, with the possible exception of $\text{Hg}_{1-x}\text{Mn}_x\text{Te}$ and $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$. Photoconductive detectors have been reported using $\text{Hg}_{1-x}\text{Mn}_x\text{Te}$, and photovoltaic detectors using both materials, but devices have not reached the maturity of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$. Also, the technology reported for these materials is very similar to that of HgCdTe, and so many of the detector design, technology and performance factors described in this chapter for HgCdTe are also relevant to other II–VI alloys.

The main purpose of this chapter is to describe the applications, device physics and technology of II–VI narrow-bandgap devices, focusing on HgCdTe. It concludes with a review of the research and development programs into third-generation infrared detector technology (so-called GEN III detectors) being performed in centers around the world.

36.0.1 Historical Perspective and Early Detectors

Passive thermal imaging is the term used to describe imaging of the natural thermal radiation emitted by all objects around us. The contrast in such images is due to temperature differences and changes in emissivity or spectral radiance of surfaces. The atmosphere

is rather inconvenient for infrared imaging and is only transparent in certain wavelength “windows”. There is a short-waveband (SW) window between 2.0 and 2.25 μm , a medium-waveband (MW) window between 3.0 and 5.0 μm and a long-waveband (LW) window between 7.5 and 14 μm . The spectral radiance curve for bodies around room temperature shows a peak that best matches the LW window and this is the preferred wavelength of operation for thermal imaging detectors. At first sight the MW window looks compromised by the photon flux, which is some two orders lower than the LW window. However, it is much easier to make detectors for this wavelength, and the photon flux disadvantage can be offset by using staring architectures that enable a longer integration (or stare) time, and so this band is very widely used. The SW window is rarely used for passive thermal imaging because of low flux, but active imaging using a source such as starlight or an infrared laser is of growing interest.

The first infrared photon detectors, based on thallous sulfide, were developed in the USA during World War I and were sensitive to about 1.4 μm . The next important developments occurred before and during World War II in Germany, with work on thin-film polycrystalline PbS devices with a response up to 2.5 μm , later extended further to the 3–5 μm region using PbSe and PbTe. During the 1950s and 1960s indium antimonide (InSb) detectors emerged that were capable of detecting wavelengths up to 5.5 μm , and during the same period impurity photoconductivity was studied in doped germanium and doped silicon. Germanium was favored at this time because it was available with fewer compensating impurities and gave better detector performance. Ge : Hg, with an impurity activation energy of 0.09 eV, provided the first practical detector to be used in real-time thermal imaging, employing linear arrays for the 8–13 μm region, although it required cooling to below 30 K. Because of the stringent cooling requirements, extrinsic silicon devices have not found favor for terrestrial applications.

The next important phase of semiconductor infrared detector research took place during the late 1960s and 1970s, when research efforts were directed towards an intrinsic detector for the 8–13 μm band that would operate more conveniently at around 80 K. Two alloy semiconductors were developed, $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (HgCdTe) and $\text{Pb}_{1-x}\text{Sn}_x\text{Te}$ (LTT). The first report of the synthesis of the semimetal HgTe and the wide-bandgap semiconductor CdTe to form the semiconductor alloy HgCdTe was published by the Royal Radar Establishment in Malvern, UK [36.1]. This landmark paper

reported both photoconductive as well as photovoltaic response at wavelengths extending out to $12\text{ }\mu\text{m}$, and pointed out that this new alloy semiconductor showed promise for intrinsic infrared detectors. Soon after many centers around the world switched detector development for major thermal-imaging programs to the HgCdTe system. High-performance LW and MW linear arrays were produced for the first generation of thermal imaging equipment developed and manufactured in the 1970s through to the present. Many tens of thousands of detectors have been delivered.

36.0.2 Introduction to HgCdTe

$\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is the material most widely used for high-performance infrared detectors at present. By changing the composition x , the detector spectral response can be made to cover the range from $1\text{ }\mu\text{m}$ to beyond $17\text{ }\mu\text{m}$. The advantages of this system arise from a number of features, notably: close lattice matching, high optical absorption coefficient, low carrier generation rate, high electron mobility and readily available doping techniques. These advantages mean that very sensitive infrared detectors can be produced at relatively high operating temperatures. HgCdTe continues to be developed as the material of choice for high-performance long-wavelength ($8\text{--}12\text{ }\mu\text{m}$) arrays and has an established market at the medium- ($3\text{--}5\text{ }\mu\text{m}$) and short-wavelength ($1\text{--}3\text{ }\mu\text{m}$) ranges.

In the LW band, the main competitive technologies are $\text{Pb}_{1-x}\text{Sn}_x\text{Te}$ (LTT) and multiple quantum well (MQW) detectors (usually using AlGaAs/GaAs technology). Work on LTT largely stopped in about the mid 1970s partly because the large dielectric constant made them unsuitable for photoconductors. However, there is more recent interest in large arrays of LTT photodiodes due to a potential cost advantage, but at present poor diode quality and excess noise makes them inferior to $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ on grounds of sensitivity alone. MQW arrays are essentially tuned to a wavelength, say $8.0\text{ }\mu\text{m}$, and need deeper cooling to suppress thermal leakage currents. The imaging performance can be good due to the absence of low-frequency noise sources but they have a much lower ultimate sensitivity than HgCdTe.

36.0.3 Introduction to Device Types

HgCdTe Photoconductive Arrays

A photoconductor usually comprises a small slab or element of material with two contacts. The aim is to detect the change in resistance of the element when the photon

flux is changed. The first HgCdTe devices were photoconductive because of the simplicity of the technology, and the relative ease of achieving near-ideal infrared performance and excellent reliability. HgCdTe photoconductive detectors have been in routine production since the early 1980s and are often called first-generation detectors. They are the key component of the US Common Module Thermal Imager and, in the form of the SPRITE detector, they feature in the UK Class II Common Module imager. Detectors for thermal imaging are most commonly fabricated with a peak response in the $8\text{--}12\text{ }\mu\text{m}$ atmospheric window region and are cooled to 80 K by means of a Joule–Thompson expansion cooler or, more recently, a Stirling engine refrigerator. They are also quite commonly made for operation in the $3\text{--}5\text{ }\mu\text{m}$ atmospheric window, and either use cooling to 80 K or employ Peltier (or thermoelectric) coolers to cool to around 200 K . Peltier-cooled MW detectors are used in many small handheld cameras and a whole host of heat sensing applications. The reliable performance, low levels of defects and easily understood physics has led to a long product life for photoconductive arrays. However the array size is limited and first-generation thermal imaging systems need to employ complex optics to scan the infrared image over the array to build up a scene. The main limitation to developing larger arrays arises from the difficulty involved with amplifying and multiplexing elements electronically on the focal plane. In consequence, there needs to be a separate electrical connection to each photoconductor so that they can be connected to low noise current amplifiers outside of the cryogenic encapsulation. For larger arrays of, say, more than several hundred elements the cryogenic encapsulations become cumbersome and expensive. Also, the power consumption can become a problem with too many elements, so this and cryogenics set a practical limit to the size of photoconductor arrays for commercial thermal imaging to a few hundred elements.

HgCdTe Photovoltaic Arrays

A photovoltaic device is essentially a light-sensitive diode. Photons absorbed in the semiconductor create electron–hole pairs, and the minority carriers diffuse to the p–n junction where they are “separated” and the voltage across the junction changes. Because there is an integration of carriers, the signal can be built up over a time called the integration time, and there is the potential for larger signals and better infrared sensitivity than can be achieved with photoconductors.

In the mid-1970s attention turned to the use of photovoltaic HgCdTe for thermal imaging applications. At

that time it was seen that, in the future, many infrared applications would need higher radiometric performance and/or higher spatial resolution than could be achieved with first-generation photoconductive infrared detectors. Photovoltaic arrays consume very little power, and can be easily multiplexed using an on-focal plane silicon chip, so they are well-suited to long linear and large, two-dimensional infrared arrays. Systems based upon such focal planes can be made smaller and lighter, with lower power consumption, and can result in much higher performance than systems based on first-generation detectors. Photovoltaic detectors can also have less low frequency noise, faster response time, and can avoid the need for complex scanning optics. To some extent these advantages are offset by the more complex processes needed to fabricate photovoltaic detectors, and so their development and industrialization have been slower, particularly for large arrays. Another point is that, unlike photoconductive detectors, the field of photovoltaic HgCdTe arrays shows a large variety of different material growth methods and device structures, often unique to individual companies and research organizations. The collective wealth of data on photovoltaic devices is therefore spread over a wide field.

Large two-dimensional and long linear arrays have the common feature that they are all mass-connected to a custom-designed silicon integrated circuit called a multiplexer or ROIC (for readout integrated circuit). The multiplexer performs the function of integrating the infrared signal and scanning the array. The evolution of large arrays was delayed by the slow emergence of silicon integrated circuits large enough to perform this

function economically. Another economic factor was matching the cost of first-generation infrared cameras which had a well-established market and had set a benchmark for display picture points and cost. Cost depends strongly on array size and infrared array pixels tend to be much larger than those commonly found in visible imaging arrays. Infrared arrays then tend to be physically large and, together with the extra complexity of the manufacturing process, relatively difficult to cost-reduce. The commercial viability of second-generation focal plane arrays has depended on matching the spatial resolution (number of elements in the array) and cost of existing first-generation systems, which was only really achieved in the 1990s.

HgCdTe Metal–Insulator–Semiconductor (MIS) Arrays

The MIS HgCdTe detector operates much like a silicon charge-coupled device (CCD). The MIS detector was the basic element in a family of “monolithic” HgCdTe arrays in which the detection, integration and multiplexing functions were all done within the HgCdTe material itself. Unlike the PV and PC detectors, the MIS device operates under strongly nonequilibrium conditions, with large electric fields in the deep depletion regions. This makes the MIS detector much more sensitive to material and process defects than the PV and PC detectors [36.2]. This sensitivity, particularly acute for LW devices, caused the monolithic approach to be abandoned in favor of various hybrid approaches where the IR photon detection is performed in HgCdTe, and the signal processing is restricted to the silicon multiplexer.

36.1 Applications and Sensor Design

Currently the most important market for HgCdTe arrays is in thermal imaging in the long (8–12 μm) and medium wavebands (3–5 μm). Often the detector is the performance-limiting component in the system, and it is necessary to use detectors with a sensitivity limited only by the random rate of arrival of photons from the scene (so-called background limited or BLIP detectors). In a narrow-gap semiconductor using both photoconductive and photovoltaic detectors, it is necessary to cool so that the thermal generation and associated excess noise are suppressed and the sensitivity becomes “BLIP-limited”. The choice of cooler is therefore a key technology in an infrared detector, and the common options are described here. One of the points to emphasize

is that the focal plane array is only one of a number of components that need to be optimized to preserve the BLIP performance. The optical design, cryogenics and signal processing are critical to maintaining the sensitivity of the focal plane array. The precision engineering involved in designing infrared detectors and coolers, particularly for survival in high shock and vibration environments, is still challenging.

The means of cooling depends on the detector type and application. For instance, photoconductive arrays often need a lot of cooling power because of the Joule heating on the focal plane and the need for many bond wires that add to the thermal load. Two-dimensional arrays, on the other hand, often consume little power



Fig. 36.1 Typical fast cooldown detector using Joule–Thompson cooling. (Courtesy of SELEX Infrared)

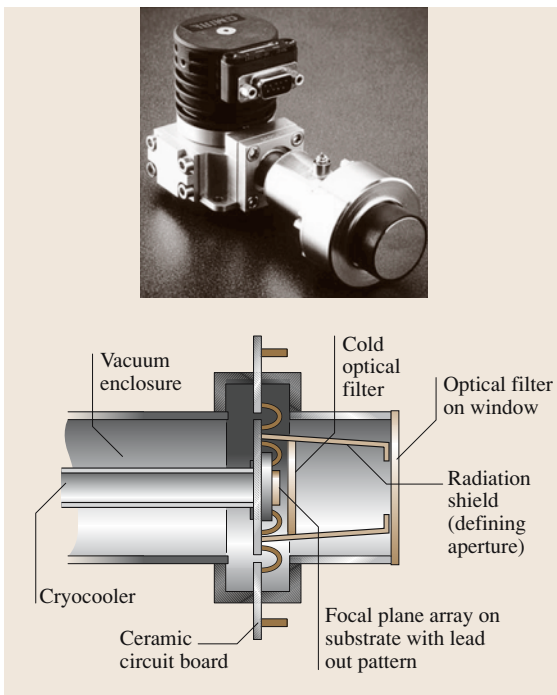


Fig. 36.2 Typical integrated detector: cooler assembly (IDCA). (Courtesy of SELEX Infrared)

and have few connections, allowing alternative cooling methods. Other considerations are cooldown time, power, weight, space-volume, robustness and life.

Historically the Joule–Thompson cooler was the first available method of cooling, relying on compressed gas bottles for power. The cooling power can be very high. Most first-generation detectors use Joule–Thompson



Fig. 36.3 Typical thermoelectrically cooled detector. (Courtesy of SELEX Infrared)

cooling. In more recent times the high cooling power has been routinely exploited in fast-cooldown applications. Figure 36.1 illustrates a typical fast-cooldown product, which can be cooled in a few seconds to liquid argon temperature (92 K). HgCdTe arrays are well-suited to this type of application because, unlike other materials such as indium antimonide or multiple quantum wells, BLIP operation is easily achieved at relatively high temperatures.

For most modern thermal imaging applications the normal means of cooling is a cryocooler based on the Stirling cycle, which can achieve temperatures as low as 55 K but is normally regulated at a temperature in the range 80–140 K. Figure 36.2 shows a typical integrated detector–cooler assembly (IDCA). The cutaway in Fig. 36.2 shows the front end of the detector in more detail. The infrared array is bonded directly on the cold finger of the cooler. To ensure only photons from the scene reach the detector, a cold stop or radiation shield is bonded to the cold finger. The cold stop will have internal baffles and special coatings to stop internal reflections and is a vital component for achieving BLIP operation of the array. The detector will have an optical pass-band filter that controls the upper and lower wavelengths. This will vary with the application, but typical bands are 3.2 to 4.2 μm , 3.7 to 4.9 μm or 7.7 to 10.6 μm . Two methods are commonly used. The filter can be incorporated in the cold stop, and is therefore cooled, producing a very efficient block to stray radiation, say from the dewar itself. Almost as good performance can be achieved with filters on the front window, and a well-designed cold stop.

The third common means of cooling is Peltier cooling (usually based on bismuth telluride devices). A typical detector is illustrated in Fig. 36.3. State of the art coolers can achieve temperatures as low as 180 K, but in general they are not as efficient as Stirling cryocoolers,

and the cooling performance can be poor in high ambient temperatures. Nevertheless the main strengths of long-term reliability, greater shock resistance and smaller weight and space-volume has created a large volume

market, especially for small photoconductor arrays. The temperature of 180 K is not quite cold enough for BLIP thermal imaging, and the applications tend to be those that can tolerate lower performance infrared detectors.

36.2 Photoconductive Detectors in HgCdTe and Related Alloys

The physics and technology of HgCdTe photoconductive detectors are described here. It also serves as an introduction to SPRITE detectors described in Sect. 36.3 and related work on other II–VI compounds summarized in Sect. 36.4.

Being quite a simple device, the photoconductor provides a convenient introduction to the physics and technology of infrared detectors made from HgCdTe. Figure 36.4 shows a photomicrograph and scanning electron microscope view of a typical photoconductive element. It comprises a small slab of material, typically 50 μm square and 8 μm thick with two ohmic contacts. The element illustrated has been “labyrinthed” to improve its performance. The device is operated using a constant current bias, supplied by means of a voltage source, and a series resistance, which is large compared to the detector resistance. The signal is measured as a voltage change across the detector, so that the device is essentially a radiation-sensitive resistor.

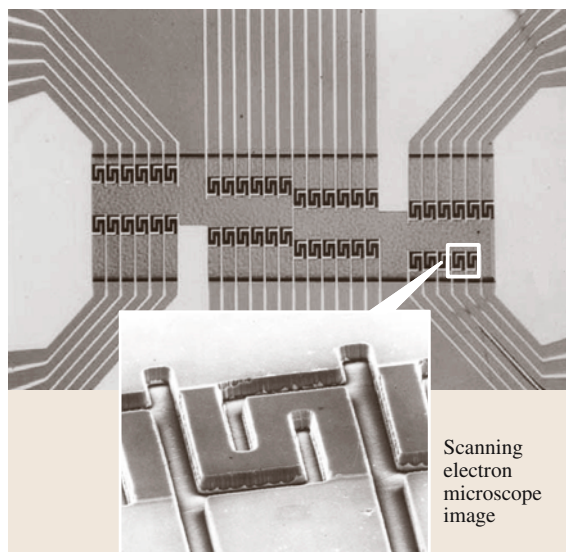


Fig. 36.4 Photomicrograph of 48-element photoconductive array using labyrinthed structures for enhanced responsivity. (Courtesy of SELEX Infrared)

36.2.1 Introduction to the Technology of Photoconductor Arrays

HgCdTe photoconductors were developed in the early 1970s when only bulk-grown material (from a Bridgman process or a solid state recrystallization process) was available. The simple device geometry allowed arrays to be manufactured by chemomechanical polishing of thick slices of material. Bulk material is still used because of the economic and convenient processes involved, but many workers [36.3–5] report arrays produced from LPE and MOVPE material. Because the electron mobility is many orders higher than the hole mobility in narrow-bandgap HgCdTe, it is advantageous to work with n-type material where the photoconductive gain is very high. n-type material can be manufactured directly, using the Bridgman growth process, or can be produced by two-temperature annealing. Chapter 14 describes the range of growth processes for HgCdTe. For photoconductors there is a requirement for a long minority carrier lifetime, which implies low levels of surface and bulk recombination centers. n-type material with a low doping level is ideal, because the fundamental carrier lifetime is more than adequate for most applications, and crystalline defects in n-type material appear to have a weak effect on the carrier lifetime. The quality of surfaces is critical because elements can be as thin as 4 μm . The technique most commonly used is to passivate the surfaces by anodic oxidation. The anodic oxide has a charge state that attracts electrons (accumulated surface) and provides an electric field which repels minority carriers, preventing them from interacting with surface recombination sites. The so-called surface recombination velocities are as low as 50 cm/s [36.6]. The individual elements are defined using a photomask and ion beam milling. As described later, ion beam milling results in a heavily doped n-type skin layer which also produces a low surface recombination velocity. Ion milling is also used to prepare the contact windows. The resulting highly doped n-type layer provides an excellent ohmic contact for electron flow, and a blocking contact to minority carrier holes. This condition enhances the minority carrier lifetime,

which leads to high signals and high sensitivity. As will be seen in the performance analysis, the device geometry and doping levels arising from the technology provide an excellent platform for producing sensitive detectors at relatively moderate cooling.

36.2.2 Theoretical Fundamentals for Long-Wavelength Arrays

Despite its apparent simplicity, a rigorous analysis of photoconductor operation is quite complex [36.7, 8]. Two-dimensional and three-dimensional treatments have been reported [36.9, 10], but the analysis here is restricted to a one-dimensional treatment.

For the purposes of simplicity, the analysis will concentrate on an n-type element, which is the most commonly used device structure. Consider a slab of HgCdTe of length l , width w and thickness t , having “ohmic” contacts formed to the end faces, and exposed to an infrared background flux of ϕ_b photons/cm².

Consider an electron–hole pair that is generated by a photon. Holes would drift towards the negative electrode with a velocity $\mu_a E$, where μ_a is the ambipolar mobility given by

$$\mu_a = (n - p)\mu_e\mu_h / (n\mu_e + p\mu_h), \quad (36.1)$$

where n and p are the electron and hole densities, and μ_e and μ_h are the electron and hole mobilities. The ambipolar mobility approximates to the minority carrier mobility in most situations, except when the temperature or the background flux is very high. At low values of the applied field, the average drift length of a minority carrier, $\mu_a E\tau$, is very much less than the detector length l , and recombination occurs predominantly in the bulk of the sample. The minority carrier density is uniform along the length of the sample (neglecting diffusion), and in the presence of background radiation, is given by:

$$p = p_0 + p_b \quad (36.2)$$

where p_0 is the thermal equilibrium concentration of holes, and

$$p_b = \eta\phi_b\tau/t \quad (36.3)$$

is the excess carrier density, in equilibrium with the background radiation, where η is the quantum efficiency.

Equations for the device parameters are given below. In large n-type detectors where the minority carrier drift length, $\mu_a E\tau$, is small compared to l , then the

responsivity in units of V/W is given by:

$$R_v = \frac{\eta V_0 \tau}{E_\lambda l w t n_0}, \quad \text{low bias} \quad (36.4)$$

where V_0 is the bias voltage and E_λ is the energy of a photon with wavelength λ .

At higher bias levels, the minority carrier lifetime in weakly doped n-type material can be made long enough to ensure that little recombination takes place before the carrier is swept to the negative contact (the so-called sweepout regime), such that $\mu_a E\tau > l$. The effective lifetime is reduced to a value τ_{eff} and, in the limit, to τ_a , where it is simply the average transit time. For simplicity we will assume that the equilibrium carrier densities are unchanged under bias. At the biases used in practical detectors, blocking contacts influence the carrier densities, and the effects must be taken into account.

The responsivity in sweepout is given by:

$$R_v(\text{max}) = \frac{\eta l}{2 E_\lambda w t n_0 \mu_h}, \quad \text{high bias} \quad (36.5)$$

Taking $\eta = 0.6$, $l = w = 50 \mu\text{m}$, $t = 8 \mu\text{m}$, $n_0 = 5 \times 10^{14} \text{ cm}^{-3}$, $E_\lambda = 0.1 \text{ eV}$ and $\mu_h = 450 \text{ cm}^2/\text{Vs}$, a value

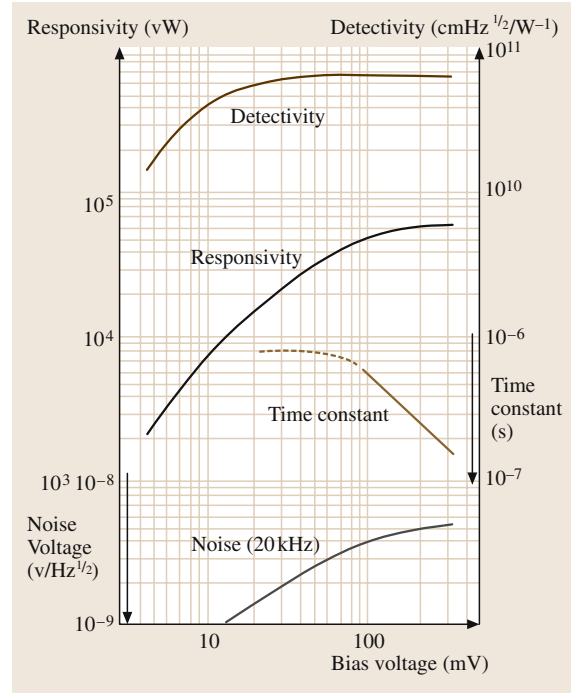


Fig. 36.5 Characteristics of a 50 μm -square HgCdTe detector, operated at 80 K, as a function of bias voltage. (After SELEX Infrared)

for the maximum responsivity of $\approx 1 \times 10^5$ V/W is obtained. The responsivity observed in practical devices is lower than the calculated value, typically by a factor of two, because the anodic oxidation technique used to produce low recombination velocity surfaces also enhances the surface conductance, shunting the bulk of the device and lowering its resistance, i.e., n_0 is effectively increased in (36.5). The behavior of a typical device, showing the saturation in responsivity with bias, is shown in Fig. 36.5.

Responsivities in photoconductors tend to be low and there have been a number of novel techniques reported to boost the signal. Several methods concentrate on prolonging the life of minority carriers. One technique [36.11, 12] is to use a remote negative contact and an opaque screen to define the sensitive area so that holes are swept under the screen and contribute to the signal for longer. Another technique [36.13, 14] is to use a low recombination velocity n^+n contact to increase the responsivity by a factor of five by blocking the recombination of holes. Even larger increases have been obtained [36.15] by using a heterojunction contact, i.e., an epitaxially grown double layer is employed with a high x region of n -type HgCdTe between the active region and the metal contacts. The theory of devices with blocking contacts has been published in several papers [36.14–16]. Significant increases in responsivity have also been obtained by increasing the element resistance by defining a “meander path” or “labyrinth” device (Fig. 36.4).

The principal noise sources are generation–recombination ($g-r$) noise due to fluctuations in the optically or thermally generated free-carrier densities and Johnson–Nyquist noise, associated with the finite resistance of the devices. In addition, noise with a spectral power density varying with frequency as $1/f$ is usually observed, which has its origin in surface trapping and mobility modulation effects.

The equation for root mean square $g-r$ noise for a small device in strong sweep-out, and in which the background-induced carrier density $p_b\tau_a/2\tau \gg p_0$, is given by:

$$V_{g-r} = \frac{1}{n_0\mu_h t} \left(\frac{l^3 \eta \phi_b B}{w} \right)^{1/2}, \quad (36.6)$$

where B is the bandwidth. Using the same parameters as used above, and assuming a background flux of 5×10^{16} cm $^{-2}$ /s ($f/2$ cold shield and 300 K ambient temperature), the value for V_{g-r} in saturation is 5×10^{-9} V/Hz $^{1/2}$. This corresponds to a noise-equivalent resistance of approximately 1.5 k Ω , and since

the actual resistance of the device is $\approx 100 \Omega$ the design of low-noise preamplifiers is relatively straightforward. A plot of the $g-r$ noise for a practical device, with a peak response wavelength of 12 μ m, exposed to 300 K background radiation in an $f/2$ field of view, is shown in Fig. 36.5. The saturation of the noise at high bias is clearly demonstrated.

A necessary condition for background-limited operation is that $p_b \gg p_0$ or

$$\frac{\eta \phi_b \tau}{t} \gg \frac{n_i^2}{n_0}, \quad (36.7)$$

where n_i is the intrinsic carrier density, which is approximately 10^{13} cm $^{-3}$. The lifetime is itself a function of the background flux, but in reduced field of view, values in excess of 2 μ s are observed. Putting $\eta = 0.6$, $t = 10 \mu$ m and $n_0 = 5 \times 10^{14}$ cm $^{-3}$, we see that background-limited operation should be observed when $\phi_b > 1 \times 10^{15}$ cm $^{-2}$ /s, i.e., a field of view greater than about 5°, and for a background scene temperature of 300 K.

It is generally the case [36.17–19] that the lifetime in good-quality n -type HgCdTe, with a cut-off wavelength near 10 μ m at 80 K, is determined by Auger 1 recombination, except possibly at very low donor concentrations near 1×10^{14} cm $^{-3}$, where Shockley–Read recombination may be significant. Auger 1 is essentially impact ionization by electrons in the high-energy tail of the Fermi–Dirac distribution, which have energies greater than the band gap E_g . An expression for Auger lifetime is:

$$\tau = 2\tau_{Ai1} \frac{n_i^2}{n(n+p)}, \quad (36.8)$$

where τ_{Ai1} is the intrinsic Auger 1 lifetime, which has a value of approximately 1×10^{-3} s.

The sensitivity of a detector is often described by the parameter called the detectivity. The detectivity (or D^*) is a signal-to-noise parameter normalized for area and bandwidth, and is used to compare photoconductor performance. Dividing the responsivity by the $g-r$ noise expression gives

$$D_\lambda^* = \frac{\eta^{1/2}}{2E_\lambda} \left(\frac{1}{\phi_b} \right). \quad (36.9)$$

The limiting D^* in zero background flux is given by

$$D_\lambda^* = \frac{\eta}{2E_\lambda} \left(\frac{2\tau_{Ai1}}{n_0 t} \right)^{1/2}. \quad (36.10)$$

Using the values for the parameters as taken earlier

$$D_\lambda^* = 1 \times 10^{12} \text{ cmHz}^{1/2}/\text{W} \quad (36.11)$$

and experimental results close to this value have been reported [36.19].

In the high-bias sweep-out condition, photoconductive detectors can have a frequency response of several MHz, the response time being determined by the transit time of the holes between the electrodes, $l^2/\mu_h V_0$, rather than the excess carrier lifetime. The effect of bias on time constant is shown for a practical device in Fig. 36.5.

Detectors for space applications, with cut-off wavelengths of between 15 and 16 μm , have been reported with D^* of $3\text{--}4 \times 10^{11} \text{ cm Hz}^{1/2}/\text{W}$, when operated at 60 K in a background flux of $2 \times 10^{15} \text{ photons/cm}^2 \text{ s}$ [36.18].

36.2.3 Special Case of Medium-Wavelength Arrays

Detectors operating in the 3–5 μm range, with an aperture of $f/1$ or faster, normally have near-BLIP performance. Figure 36.6 shows the D^* versus cut-off wavelength for n-type photoconductive detectors. The theoretical maxima include Auger generation and radiative generation only. The experimental points are for 230 μm square n-type detectors, except where indicated. Small detectors in very low background flux conditions are expected to be Johnson noise-limited due to sweep-out. The limiting value of detectivity for n-type detectors in zero background flux [36.20, 21] is

$$D^*_{\lambda} = \frac{\eta l}{4 E_{\lambda} t^{1/2}} \left(\frac{q}{kT} \right)^{1/2} \left(\frac{b}{\mu_h n} \right)^{1/2} \cong 1.7 \times 10^{19} \left(\frac{b}{\mu_h n} \right)^{1/2}, \quad (36.12)$$

where the numerical value is obtained assuming $l = w = 100 \mu\text{m}$, $t = 10 \mu\text{m}$, $\eta = 0.7$ and $E_{\lambda} = 0.25 \text{ eV}$.

36.2.4 Nonequilibrium Effects in Photoconductors

So far, the modeling of photoconductive devices has assumed equilibrium conditions for carrier densities. In reality, the application of bias fields results in departures from thermal equilibrium. An example occurs at the positive contact, where the highly doped buffer layer results in negligible hole emission (so-called excluding contact). When the bias is applied, no holes can be in-

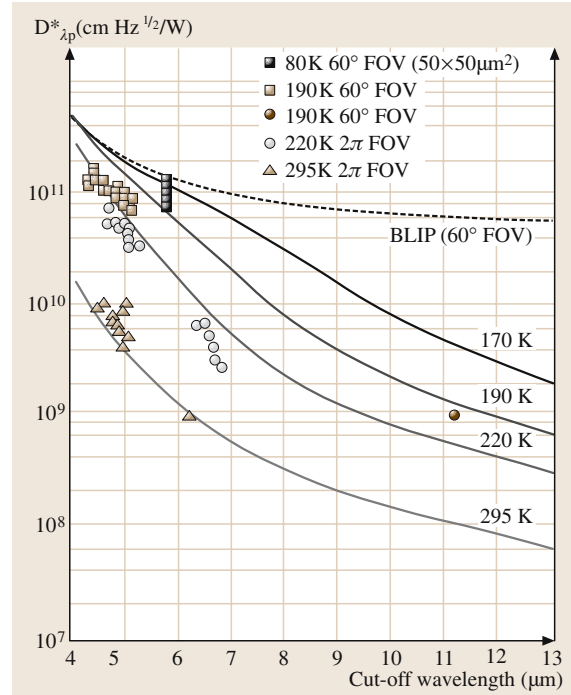


Fig. 36.6 D^*_{λ} versus cut-off wavelength for n-type photoconductive detectors. *Solid square* 80 K 60° FOV ($50 \times 50 \mu\text{m}^2$); *open square* 190 K 80° FOV; *solid circle* 190 K 60° FOV; *open circle* 220 K 2π FOV; *open triangle* 295 K 2π FOV; *solid line* Theoretical maximum in equilibrium (Auger 1, Auger 7 and radiative generation); *dotted line* BLIP (60° FOV, 30% reflection loss). (After SELEX Infrared, with permission)

jected into the material to sustain the flow of holes in the bulk away from the contact. As a consequence, p is reduced for some distance away from the contact. In order to maintain near space-charge neutrality, n also falls to a value close to the extrinsic value. The dominant thermal generation process in n-type material is Auger 1. For the example shown, n is reduced by a factor of about thirty, which would reduce the Auger 1 generation rate by the same factor, and increase the detectivity by a factor of more than five if other noise sources, such as Shockley–Read, were low. In devices with noninjecting contacts this effect results in a significant operating temperature advantage.

36.3 SPRITE Detectors

The SPRITE detector was invented in the early 1970s [36.20, 22, 23] and it forms the basis of the UK common module detector, produced in large numbers. A thorough description of the device is reported in [36.24]. It is a very elegant method of amplifying the sensitivity of a photoconductive device. It is also a very compact device, which enables highly efficient radiation shielding and economical use of material. SPRITE detectors are used in serial-parallel scan imaging systems, because they perform a time-delay and signal integration function, in addition to a detector function. The operating principle of the device is illustrated in Figs. 36.7a and 36.7b. It consists of a strip of n-type HgCdTe, typically 700 μm long, 60 μm wide and 10 μm thick, with three ohmic contacts. A constant-current circuit provides bias through the two end contacts. The third contact is a potential probe for signal read-out. There are two important points concerning the operation of the device. Firstly, the bias field must be sufficiently high to ensure that excess minority carriers, generated along the filament length, can reach the negative end contact before recombination occurs. Secondly, the bias field must be chosen so that the ambipolar drift velocity V_a is equal to the image scan velocity. SPRITE detectors have been commonly used in 8-row, 16-row and 24-row arrays and are commercially available in these forms.

Consider now an image feature scanned along the filament, as illustrated in Fig. 36.7a. The density of excess carriers in the filament, at a position corresponding to the illuminated element, increases during the scan, as illustrated in Fig. 36.7b. The carrier concentration increases during the scan as $[1 - \exp(-x/v_s\tau)]$, where x is the distance from the positive bias contact along the scan direction, and τ is the excess carrier lifetime. When the illuminated region enters the readout zone, the increased conductivity modulates the voltage on the contact and provides an output signal. Since the integration time of the signal flux, which is τ in a long element (or L/v_a in a shorter element), is greater than the dwell time on a conventional discrete element in a fast-scanned, serial system, a larger conductivity modulation and, hence, a larger output signal is observed.

Therefore, the SPRITE detector continuously performs a time-delay-integration function and the signal-to-noise ratio is improved as a result of coherent integration of the signal and incoherent integration of the noise. The net gain in the signal-to-noise ratio is proportional to the square root of the integration time. For an amplifier-limited or a Johnson noise-limited detector, the gain is proportional to the integration time.

The viability of the device depends on three aspects. The minority carrier lifetime must be maximized to give the longest possible integration length, the speed of the carriers (holes) must match a practical image scanning speed and the thermal spread of the carriers must be minimized as this will blur the image. In practice, this implies the need for large τ values to provide long integration times, and low minority carrier diffusion lengths Q_h to provide good spatial resolution. The material properties obtained for low carrier concentration n-type HgCdTe (as shown in Table 36.1) proves this to be a very suitable material for the device. Lifetime values of about 2 μs are obtained in alloy compositions suitable for 8–14 μm , 80 K operation, with an extrinsic carrier density of $\approx 10^{14} \text{ cm}^{-3}$. Recombination is principally by the Auger 1 process in these compositions, and longer

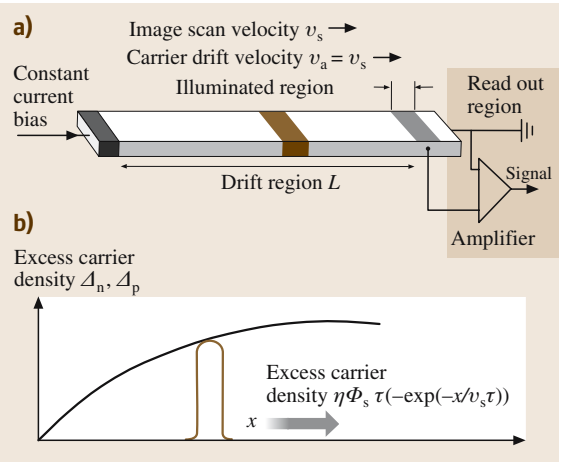


Fig. 36.7 The operating principle of a SPRITE detector

Table 36.1 Properties of n-type HgCdTe

Waveband (μm)	Operating temperature (K)	Lifetime (μs)	Hole mobility (μ_h) ($\text{cm}^2/\text{V/s}$)	D_h (cm^2/s)	Diffusion length Q_h (μm)
8–14	80	2–5	480	3.2	25
3–5	230	15–30	100	2.0	5

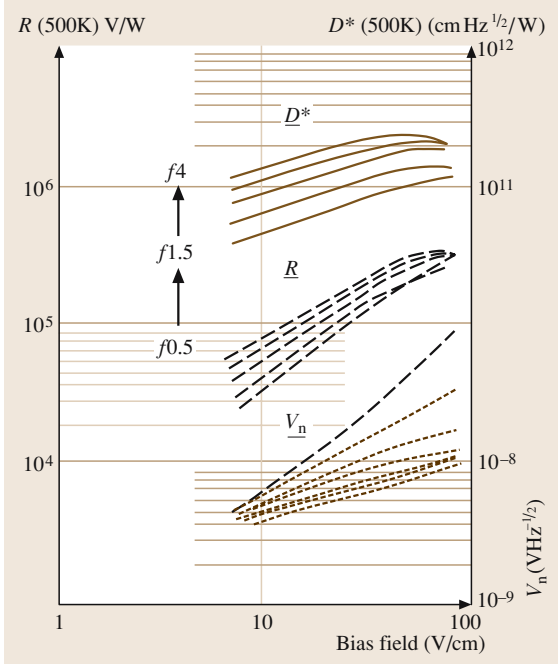


Fig. 36.8 Typical performance of a LW SPRITE detector ($\lambda_c = 11 \mu\text{m}$, $T = 77 \text{ K}$). (After [36.25])

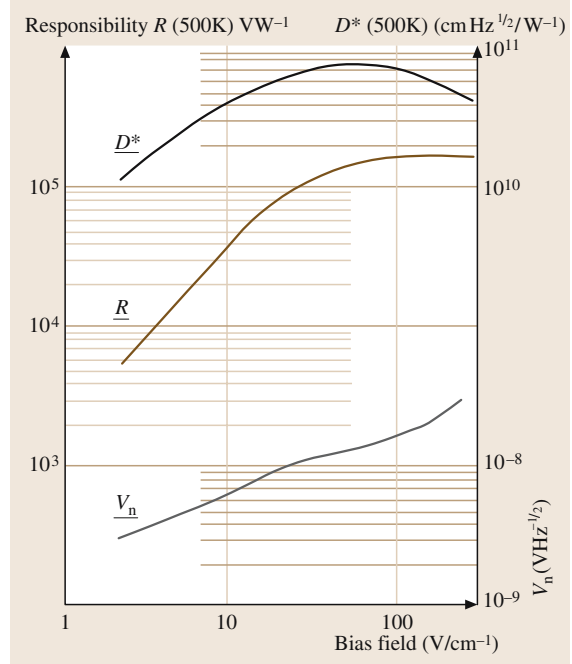


Fig. 36.9 Typical performance of a 3–5 μm SPRITE operating at 190 K. (After [36.25])

lifetimes, up to $\approx 10 \mu\text{s}$ (achievable in low-background conditions), may be obtained from lower carrier density material. Lifetime values of 15–30 μs are obtained in compositions used for 3–5 μm operation at thermoelectric temperatures. The low hole mobility in HgCdTe results in short diffusion lengths, even with long lifetimes, yielding a spatial resolution in the 8–14 μm band SPRITE devices of $\approx 50 \mu\text{m}$ and $\approx 100 \mu\text{m}$ in 3–5 μm devices if the full integration is used.

The detectivity for a long, background-limited device in which $L \gg \mu_a E \tau$ and $\eta \phi_b \tau / t \gg p_0$ is

$$D_{\lambda}^* = \frac{\eta^{1/2}}{2E_{\lambda}} \left(\frac{1}{\phi_b w} \right)^{1/2} \left\{ 1 - \frac{\tau}{\tau_a} [1 - \exp(\tau_a/\tau)] \right\}^{1/2}, \quad (36.13)$$

where w is the element width, l is the length of the readout region, and $\tau_a = \ell/\nu_a$ is the transit time through the readout region. At sufficiently high scan speeds, such that $\tau_a \ll \tau$,

$$D_{\lambda}^* = (2\eta)^{1/2} D^* (\text{BLIP}) \left(\frac{l}{w} \right)^{1/2} \left(\frac{\tau}{\tau_a} \right)^{1/2}. \quad (36.14)$$

It is convenient to express D_{λ}^* in terms of the pixel rate S which, for a nominal resolution size of $w \times w$, is ν_s/w .

Thus, in the high scan speed limit

$$D_{\lambda}^* = (2\eta)^{1/2} D^* (\text{BLIP}) (S\tau)^{1/2}. \quad (36.15)$$

The number of background-limited elements that would be required to provide the same performance in a serial array is

$$N_{\text{eq}} (\text{BLIP}) = 2S\tau. \quad (36.16)$$

For example, a 60 μm wide element scanned at a speed of $2 \times 10^4 \text{ cm/s}$, and with τ equal to 2 μs , gives $N_{\text{eq}} (\text{BLIP}) = 13$.

The performance achieved in the 8–12 μm band is illustrated in Fig. 36.8. It may be seen that very high values of responsivity are obtained, as well as high values of D^* . The latter increases with the square root of the bias field, except at the highest fields, where Joule heating raises the element temperature. The detectivity increases with increasing cold-shield effective f /number to about $f/4$. It is, in practice, desirable to operate with efficient cold shielding with an f /number of ≥ 2 to avoid reductions of the carrier lifetime resulting from increased carrier density due to the background flux. An example of the results obtained from a SPRITE operating in the 3–5 μm band is shown in Fig. 36.9. Useful performance

in this band can be obtained at temperatures up to about 240 K.

The spatial resolution of the SPRITE detector when the scan velocity and the carrier velocity are matched throughout the device length, is determined by the diffusive spread of the photogenerated carriers and the spatial averaging in the readout zone. This may be expressed through the modulation transfer function (MTF):

$$\text{MTF} = \left(\frac{1}{1 + k_s^2 Q_h^3} \right) \left[\frac{2 \sin(k_s l / 2)}{k_s l} \right], \quad (36.17)$$

where k_s is the spatial frequency. The behavior of noise in the device at high frequencies is described in several papers [36.26,27]. Some optimization of the shape of the filament has been carried out, involving a tapering of the read-out zone and a slight tapering of the main body of

the element [36.28]. Following these modifications, very good agreement with (36.17) was obtained. In practice, the length of the read-out zone can be chosen sufficiently small that the diffusion term sets the limit. It was believed initially that this limit was fundamental, and that the only way in which it could be reduced was to restrict the integration by means of a shorter device. In fact, several different methods have been discussed [36.29] by means of which better spatial resolution can be obtained at the expense of additional power dissipation on the focal plane. The technique that has found most favor with system designers has been to increase the focal length of the detector lens in an afocal system, usually by employing anamorphic optics [36.30]. This allows the diffusion spot radius to be reduced below the diffraction spot radius.

36.4 Photoconductive Detectors in Closely Related Alloys

There has been some interest in developing alternative ternary alloys to replace HgCdTe, as from theoretical considerations it has been shown that the already weak Hg–Te lattice bond is further destabilized by alloying with CdTe [36.31]. It is predicted that the Hg–Te bond may be more stable in alternative alloys, leading to materials with increased hardness and detectors with better temperature stability. Apart from Cd and Zn, the other elements capable of opening up a bandgap in the semimetals HgTe and HgSe are Mn and Mg. In the case of Mg, little reinforcement of the Hg–Te bond occurs at compositions appropriate to 10 μm detectors, and in the case of $\text{Hg}_{1-x}\text{Cd}_x\text{Se}$, difficulties have been experienced in obtaining type-conversion and lightly doped n-type material. Most of the attention, therefore, has been focused on $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ and $\text{Hg}_{1-x}\text{Mn}_x\text{Te}$. As mentioned in the introduction, the technology and general properties of these materials are very similar to HgCdTe and much of the device processing is common to both. The status of the research is described by [36.32].

The hardness of $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ is about a factor of two better than HgCdTe and interdiffusion about one tenth, so this compound offers promise of competing with HgCdTe. However $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ presents more difficult material problems than HgCdTe. For instance, the separation of the liquidus and solidus curves is large and leads to high segregation coefficients. Also, the solidus lines are flat, leading to a strong composition dependence on growth temperature, and a very high mercury

overpressure is needed for bulk crystal growth. The best crystals have been grown by the traveling heater method (THM). A D_λ^* value of $8 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$ has been quoted [36.33] for photoconductive detectors at 80 K prepared from LPE material, a value substantially lower than that typical of HgCdTe devices. 10.6 μm laser detectors with D_λ^* values in the range $3\text{--}6.5 \times 10^7 \text{ cm Hz}^{1/2}/\text{W}$ have been reported [36.34], comparable to those observed in equilibrium devices in HgCdTe. The most promising results, however, are those for long-wavelength detectors, of interest for space applications, with cut-off wavelengths as long as 17 μm [36.35]. The D_λ^* values were $8\text{--}10 \times 10^{10} \text{ cm Hz}^{1/2}/\text{W}$, measured at 65 K. Photodiodes have been produced by all of the established techniques, including ion implantation into THM material [36.36] and Hg diffusion into isothermal vapor phase material [36.37]. Both report comparable performance to HgCdTe diodes. The maximum annealing temperature is reported to be 10 to 20° better than HgCdTe. Challenges for the future will be to match the uniformity and defect levels of HgCdTe.

$\text{Hg}_{1-x}\text{Mn}_x\text{Te}$ is a semimagnetic narrow-gap semiconductor, but strictly it is not a II–VI ternary compound because Mn is not a group II element but is included here for completeness. $\text{Hg}_{1-x}\text{Mn}_x\text{Te}$ also has material disadvantages compared to HgCdTe and HgMnTe crystals need to be much more uniform in composition to achieve the same wavelength uniformity. For epitaxial growth the strong variation of lattice parameter with composition is also thought to be a disadvantage com-

pared with HgCdTe. Photodiodes have been produced by Hg diffusion into THM material and VPE material

grown on CdMnTe substrates, but the RoA values were low [36.38] and not yet competitive with other materials.

36.5 Conclusions on Photoconductive HgCdTe Detectors

Photoconductive HgCdTe detectors have been very successful in producing arrays of up to a few hundred elements or the equivalent in SPRITEs, for use in first-generation thermal imaging systems. The fundamental properties of HgCdTe have been found to be near-ideal for the fabrication of high-responsivity single elements, and for the fabrication of SPRITEs with very high responsivity and detectivity together with good spatial resolution. Frequency response into the MHz region is obtained from both types of device due to sweep-out effects. The limitations of photoconductive detectors are apparent when very large focal

plane arrays are required. The low impedance of the photoconductor makes it unsuitable for coupling to direct injection gates of silicon charge transfer devices or MOS field effect transistors, therefore each element requires a lead-out through the vacuum encapsulation to an off-focal plane amplifier. The complexity of the dewar, therefore, limits the size of the array. This problem is considerably reduced in the SPRITE, but in this case, the effect of the Joule heating in the elements on the heat load of the cooler limits the maximum array size. For comprehensive reviews see [36.8, 39–41].

36.6 Photovoltaic Devices in HgCdTe

Photovoltaic arrays have inherently low power consumption and can be easily connected to a silicon integrated circuit to produce a retina-like focal plane array. Such arrays are key to so-called second-generation thermal imaging cameras which break through the performance limits imposed by photoconductive arrays in first-generation cameras. Photovoltaic devices tend to fall into two categories: long linear arrays of long-wavelength (8–12 μm) diodes and matrix arrays of medium-wavelength (3–5 μm) diodes.

Long linear arrays are required for systems with just one dimension of scanning. Here HgCdTe is the sensor material of choice because the stringent sensitivity requirements of these systems demands a high quantum efficiency and near BLIP-limited performance detectors. Development work in the 1980s and 1990s in the USA and Europe has led to a family of detectors for long-waveband thermal imaging, such as the US SADA II system based on a 480×6 format or the UK STAIRS “C” system based on a 768×8 array format (producing a high-definition 1250×768 image). In the detector examples quoted there are six or eight rows of diodes and the signal from each row is delayed in time and added to provide an enhancement. This is called time delay and integration (TDI). An important advantage of this scheme is that defective pixels can be deselected, so typically the STAIRS system will only use the best six of

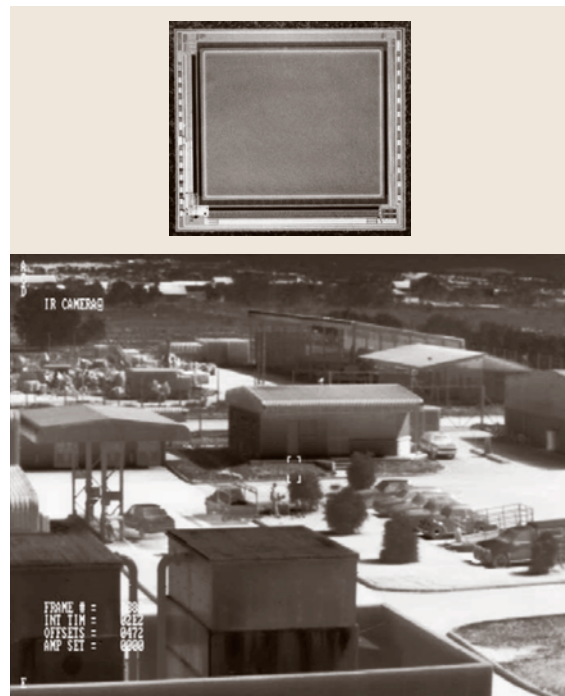


Fig. 36.10 Medium-wavelength infrared image from a 384×288 element array (OSPREY) and the Kenis System. (Courtesy Denel Ltd)

the eight diodes in one channel. Consequently the detector performance can be relatively insensitive to defects in the HgCdTe, and this is the reason why LW long linear arrays have reached maturity much earlier than LW matrix arrays.

Long linear arrays can also use the MW band, but the flux levels are over an order less than the LW band and the performance is compromised by the short integration time in a scanned system. MW detectors then tend to use matrix arrays where the integration time can approach the frame time. The problem of point defects is considerably reduced in the medium waveband and large 2-D arrays are more practical. The emphasis for MW detectors is on developing wafer-scale processes to provide large, economical 2-D arrays for staring thermal imaging cameras. The MW band thermal imaging market is concentrated on arrays using half-TV (320×256 or 384×288) or full-TV formats (640×480 or 640×512), and is divided between indium antimonide (InSb), HgCdTe and platinum silicide. Broadly, manufacturing companies from the USA and Israel tend to prefer InSb, and European and some US companies have specialized in HgCdTe. Currently, infrared cameras based on HgCdTe arrays are in production and are producing remarkable sensitivities, over an order better than first-generation cameras. Figure 36.10 shows a representative image from a state of the art HgCdTe 2-D array.

Unlike photoconductive arrays, manufacturers use a variety of different technologies for photovoltaic arrays. Sections 36.6.1 to 36.6.3 describe the theory of photovoltaic detectors and the fundamental principles behind HgCdTe device technology. These principles help to explain the various approaches used by manufacturers to produce detectors in both the LW and MW bands described in Sects. 36.6.4 to 36.6.7. Section 36.6.8 describes the research and development progress in so-called Gen III (third-generation thermal imaging equipment) programs aimed at advanced infrared detectors.

36.6.1 Ideal Photovoltaic Devices

The current–voltage characteristic for an ideal diffusion-limited diode exposed to a photon flux ϕ_λ is given by:

$$I_d = I_s \left(\exp \frac{qV_d}{kT} - 1 \right) - qA\eta\phi_\lambda, \quad (36.18)$$

where I_d is the diode current, I_s is the diffusion current, V_d is the diode bias voltage (taking negative values in reverse-bias), η is the quantum efficiency and A is

the detector area. The second term in (36.18) is the photocurrent, I_{ph} . The quantum efficiency is given by:

$$\eta = (1 - R)[1 - \exp(-\alpha_\lambda t)]F, \quad (36.19)$$

where R is the reflectivity of the front surface, normally minimized by an antireflection coating, α_λ is the absorption coefficient, t is the sample thickness, which is typically around $6\text{--}10\ \mu\text{m}$ to give adequate absorption in HgCdTe detectors, and F is a geometry factor which describes the number of photogenerated carriers within the pixel which reach the junction before recombining.

The condition that the detector be background limited is crucial for high-performance systems, and to meet this condition the internal thermal generation must be much less than the photon generation in the lowest flux case, i.e. $I_s \ll I_{ph}$. By differentiating (36.18), I_s is given by kT/qR_0 , where R_0 is the zero bias resistance. The condition for background-limited performance is therefore often written as $R_0 \gg kT/qI_{ph}$, and this is the origin of the commonly used figure of merit, R_0A . As an example, in the case of a long waveband system, say $10\ \mu\text{m}$ cut-off, with an F2 optic and a lowest background scene temperature of -40°C , an R_0A of greater than $30\ \Omega\text{cm}^2$ is required. Detectors need to be cooled sufficiently to suppress thermally generated currents, but there are always pressures to avoid very low temperatures to give savings in power and cooldown time and better engine reliability. This is particularly the case for cut-off wavelengths longer than $10\ \mu\text{m}$, and detector technologies have emerged that provide high R_0A as a first priority.

The white noise current at low frequencies is given by:

$$i_n^2 = 2qI_s \left[1 + \exp \left(\frac{qV_d}{kT} \right) \right] B + 2q(qA\eta\phi_\lambda)B, \quad (36.20)$$

where B is the bandwidth in Hz. In the absence of photocurrent, the noise at zero bias is equal to the Johnson noise, $4kTB/R_0$. In reverse bias it tends to the normal expression for shot noise, $2q(I_s + I_{ph})B$. Note that the mean-square shot noise in reverse bias is half that of the Johnson noise at zero bias.

The current responsivity is $\eta q/E_\lambda$, where E_λ is the photon energy, and this leads to the general expression for detectivity:

$$D_\lambda^* = \frac{\eta q}{E_\lambda} \sqrt{\frac{A}{i_n^2}}. \quad (36.21)$$

When the photocurrent exceeds the diffusion current, the device is said to be background limited (BLIP), and

the first term in (36.20) can be ignored. The background-limited detectivity is then given by:

$$D_{\lambda}^* (\text{BLIP}) = \frac{1}{E_{\lambda}} \sqrt{\frac{\eta}{2\phi_{\lambda}}} \quad (36.22)$$

Note that, in comparison with (36.9), the background-limited detectivity of a photodiode is a factor of $\sqrt{2}$ better than that of a photoconductor.

36.6.2 Nonideal Behavior in HgCdTe Diodes

In medium-waveband arrays the current–voltage (I–V) characteristic is usually close to ideal, but at longer wavelength a number of leakage currents can impact on the I–V characteristics and degrade the performance of the detector. Figure 36.11 illustrates a typical I–V characteristic for a LW diode, along with two types of HgCdTe photodiode, a planar diode and a via-hole diode. The diagram illustrates how factors such as thermally generated current, shunt resistance and breakdown currents at high bias can effect the diode behavior. The performance can be degraded directly; for exam-

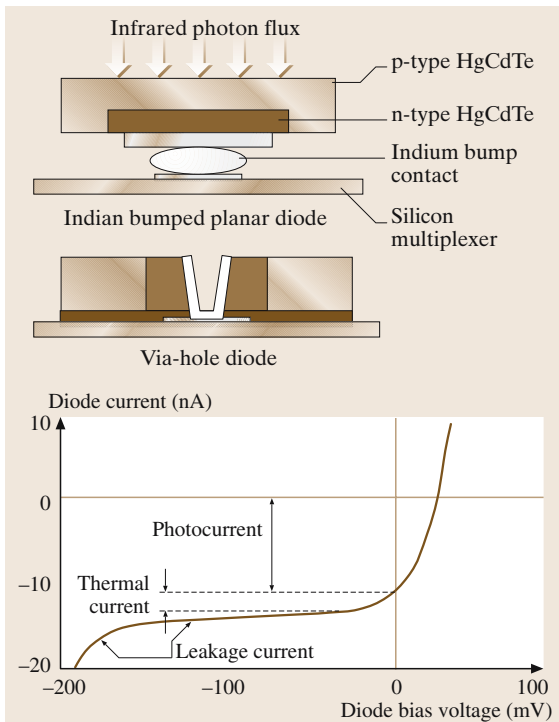


Fig. 36.11 Types of HgCdTe diode and a typical current–voltage characteristic for a LW diode showing main nonideal features

ple, extra currents produce white noise which may be significant if the leakage currents are of the same order as the photocurrent. In addition, some leakage current mechanisms are associated with trapping or tunneling and often produce excess, low-frequency noise (often called $1/f$ noise due to its typical trend with frequency). However an additional constraint arises from the need to inject photocurrent into a silicon multiplexer. The simplest method is called direct injection and literally means injection of the photocurrent into a common gate MOSFET. A capacitor on the drain performs the integration. Good injection efficiency is only achieved if the diode has a dynamic resistance higher than the input impedance of the MOSFET. Unless the dynamic impedance of the diode is high enough, the signal can be attenuated, additional noise added and nonlinearity generated in the response. Electronic correction for nonuniformity in the array is essential in IR detectors and nonlinearity can add spatial noise to the image after nonuniformity correction.

Achieving a high dynamic resistance has been a driver in the development of most infrared technologies. Very roughly there is a distinction between medium-wavelength (MW) arrays with cut-offs up to $6\text{ }\mu\text{m}$ and long-wavelength (LW) arrays. In MW arrays high injection efficiencies are routinely seen because the photodiodes are near-ideal. In the long waveband there will be a limit to the upper cut-off wavelength and/or flux levels that routinely allow high injection efficiencies using direct injection.

In order to inject the photocurrent signal efficiently into the input MOSFET, the input impedance of the MOSFET must be much lower than the internal dynamic resistance of the photodiode. The input impedance of a MOSFET is a function of the source-drain current (in this case the total diode current) and is usually expressed in terms of the transconductance, g_m , given by $qI_d/(nKT)$ for low injected currents (n is an ideality factor that can vary with technology usually in the range 1–2). The expression for injection efficiency ε given approximately by:

$$\varepsilon = \frac{R_V I_d}{R_V I_d + \frac{nKT}{q}} \quad (36.23)$$

where R_V is the dynamic impedance of the diode and I_d is the total injected diode current equal to I_{ph} in the background-limited case.

In the background-limited case, (36.23) becomes a simple function of $I_{ph} R_V$. For a very high performance system requiring an efficiency of 99.9%, the value of $I_{ph} R_V$ approaches 10 V, and in systems with slow

optics and low background temperatures, the requirement for R_v can exceed $10^9 \Omega$. In the long waveband this can present a technological challenge. There are more complex injection circuits that effectively reduce the input impedance and allow lower HgCdTe resistances to be used, but these require more silicon area and a higher power consumption, which may not be practical. The choice of buffer circuit depends on the application, and good comprehensive summaries are presented by [36.42, 43].

For most applications the detector performance depends on operating the diode in a small reverse bias where the dynamic resistance is at a maximum. It is then necessary to minimize extraneous leakage currents. The control of these leakage currents and the associated low-frequency noise is therefore of crucial interest to the device engineer. The origin of the leakage currents and the measures that are taken to minimize them are explained here in detail.

36.6.3 Theoretical Foundations of HgCdTe Array Technology

Good detailed summaries of photodiode fundamentals have been provided by [36.44, 45] and the reference lists contained therein. The purpose of this section is to highlight the key issues that control the thermal generation, leakage currents and quantum efficiency in detectors. From this analysis it is easier to understand the measures taken by detector manufacturers to make practical high-performance detectors.

Thermal Diffusion Currents in HgCdTe

Expressions for the diffusion current in photovoltaic devices have been derived [36.44] and the fundamental R_0A expression is given in (36.24):

$$R_0A(n - \text{side}) = \frac{N_d}{qn_i^2} \sqrt{\frac{kT\tau_h}{q\mu_h}}, \quad (36.24)$$

where N_d is the net donor concentration on the n-side, n_i is the intrinsic carrier concentration, which dominates the temperature dependence, and τ_h and μ_h are the minority carrier lifetime and mobility, respectively.

A similar expression describes the contribution of the p-side. In practice, the diffusion length L in normally doped HgCdTe is often larger than the 6–10 μm needed for effective infrared radiation absorption. In this case, the volume available for the generation of diffusion current is restricted, and a suitable modified expression

is:

$$R_0A(n - \text{side}) = \frac{kTN_d\tau_h}{q^2n_i^2\tau_n}, \quad (36.25)$$

where τ_n is the thickness of the n-type material.

In the case of L being greater than t_n , the surfaces and contacts can act as sources of extra diffusion current if the surface recombination velocity is greater than the diffusion velocity D/L . It is essential to ensure properly passivated surfaces and for contacts employ a minority carrier barrier, such as a higher doped or wider band gap layer under the metal contact.

The behavior of (36.25) with doping depends upon the dominant recombination process, i.e. radiative, Auger or Shockley-Read (S-R). Auger recombination in HgCdTe is a phenomenon involving the interaction of three carriers. The Auger 1 lifetime in n-type material is due to the interaction of two electrons and a hole and is generally minimized by using a low carrier concentration on the n-side. Fortunately it is relatively easy to produce low carrier concentrations in n-type material ($< 5 \times 10^{14} \text{ cm}^{-3}$).

In p-type material Auger recombination involves two holes and an electron and is referred to as Auger 7. The lifetime in p-type HgCdTe is reported [36.46–48] to show an inverse, linear dependence on doping, and this is attributed to the Shockley-Read process. S-R recombination is often associated with the Hg vacancy and can be modeled by a strong donor level located $\approx 30 \text{ mV}$ from the conduction band, which appears to be independent of composition in the $x = 0.2$ to 0.3 range. The density of these donor S-R centers is said to be proportional to the Hg vacancy concentration, but lower by a factor of ≈ 20 . If Hg vacancies are replaced by acceptor atoms such as Cu, Ag, Na or Au, the lifetime can be increased by over an order of magnitude. The use of extrinsic doping is an effective way to reduce thermal currents from the p-side in homojunction devices.

In order to engineer a detector with low thermal diffusion current (or high R_0A) it is important to recognize that it is only necessary to use one side of the junction for collecting photocurrent; the other side, in principle, can be made with a wider bandgap, thereby minimizing the thermal current contribution. Devices with layers of different band gaps are called heterostructures. The most common design is to use a wider bandgap on the p-side to reduce the thermal diffusion current and use the n-side as an absorber taking advantage of the long minority carrier diffusion lengths to maximize the quantum efficiency.

Thermal Generation Through Traps in the Depletion Region

The thermal generation rate g_{dep} within the depletion region via traps is given by the usual Shockley-Read expression:

$$g_{\text{dep}} = n_i^2 (n_1 \tau_{n0} + p_1 \tau_{p0})^{-1}, \quad (36.26)$$

where n_1 and p_1 are the electron and hole concentrations which would be obtained if the Fermi energy was at the trap energy, and τ_{n0} and τ_{p0} are the lifetimes in the strongly n-type and p-type regions.

The leakage current is often known as g–r current. Normally, one of the terms in the denominator of (36.26) will dominate, and for the case of a trap at the intrinsic level, n_1 and $p_1 = n_i$, giving $g_{\text{dep}} = n_i/\tau$. It is the weaker dependence on n_i , and therefore on temperature, that distinguishes generation within the depletion layer from thermal diffusion current.

Where the depletion region intercepts the surface, there is often enhanced generation due to the presence of a high density of interface states. This can be exacerbated if the surface passivation is not properly optimized to give a flat band potential at the junction. In an extreme case, the surface on one side of the junction may become inverted, creating an extension of the depletion layer along the surface, and leading to high generation currents. A practical solution to this mechanism is to widen the band gap in the material where the junction intercepts the surface, so-called heteropassivation. This can be achieved by using a thin film of CdTe together with a low-temperature anneal [36.49], and this is the commonest passivation technique used by manufacturers.

Interband Tunneling

Due to the very low effective mass of the electron in HgCdTe, direct band-to-band tunneling can occur from filled states in the valence band to empty states at the same energy in the conduction band. An expression for the current due to this process has been developed by [36.50]. The tunneling current increases very rapidly as the applied voltage or doping is increased, or the band gap or temperature is decreased. It is the normal modern practice to use low doping on one side of the junction to minimize interband tunneling under normal operating conditions, but nevertheless, at low temperature or for very long wavelength devices, interband tunneling can become dominant.

Trap-Assisted Tunneling

Trap-assisted tunneling is generally accepted to be one of the main causes of leakage current and excess noise in

LW diode arrays, but it is not a fundamental limitation, often being associated with impurities or structural defects within the depletion region. The definition of a trap is a center with a capture coefficient for minority carriers many times larger than that for majority carriers (otherwise it is a generation center). The role played by traps in the depletion region is very complicated, allowing for three possible two-step processes: thermal-tunnel, tunnel-thermal and tunnel-tunnel. The nonthermal step can include tunneling of electrons from the valence band to traps, and tunneling from the traps to the conduction band.

The formulation for trap-assisted tunneling in HgCdTe has been developed [36.51, 52] based upon original work in silicon [36.53]. Models based on a thermal-tunnel process can explain the bias-dependent behavior of reverse current in long-wavelength diodes [36.54–56].

A trap-assisted tunneling process can explain the observed behavior of n⁺–p diodes [36.57]. The process involves a thermally excited, bulk Shockley-Read center, modeled for the special case of the trap residing at the Fermi level. The physical picture for this assumption is that there is a uniform distribution of trapping centers throughout the bandgap and the barrier for tunneling is lowest at the uppermost center that is still occupied. Hence, the occupied trapping center that coincides with the Fermi level has the highest transition probability and plays the dominant role in the thermal trap-assisted tunneling process. This model helps to explain several unusual properties, such as an observed reduction in leakage current when the temperature is reduced or the doping level increased. Note that this is in contrast to direct band-to-band tunneling. Also, there is a much weaker dependence on diode bias voltage. The model also predicts the commonly observed “ohmic” region illustrated in Fig. 36.11 at a bias of around –100 mV.

The physical origin of the trap has not been established yet but it appears to be an acceptor-type impurity or defect within the depletion region. The population of such traps is likely to be dependent on the HgCdTe material and the junction forming technology and so the quality of LW diodes is highly process-specific.

Impact Ionization

Underlying trap-assisted tunneling, a more fundamental source of leakage current has been proposed, which is called impact ionization.

The reverse bias characteristics of homojunction arrays often show behavior that is not easily explained by conventional trapping mechanisms. For instance, the

product of p-side diffusion currents (including photocurrent) and reverse bias resistance is often observed to be insensitive to temperature and cut-off wavelength over a wide range (at least 4–11 $\mu\text{m s}$), and the current increases much more slowly with reverse bias than tunneling models would predict. A model based upon an impact ionization effect within the depletion layer gives a good fit to these observations [36.21]. The effect arises because in HgCdTe the electron scattering mechanisms tend to be weak and hot electrons can penetrate deep into the conduction band where they readily avalanche. Leakage current arises because extra electron-hole pairs are created within the depletion region due to impact ionization by minority carrier electrons from the p-side. The leakage mechanism has been confirmed to have a linear relationship with optically injected minority carriers over a wide temperature range [36.58].

Calculations have been performed on the effect of impact ionization on homojunction performance as a function of the doping levels [36.59]. These predict that to achieve a high dynamic resistance and therefore a high injection efficiency, the n-side doping must be very low. Routinely achieving low carrier concentrations ($< 5 \times 10^{14} \text{ cm}^{-3}$) is an important aim for homojunction technology. Most manufacturers introduce a donor, such as indium, to the crystal to control the n-type level to around $3\text{--}5 \times 10^{14} \text{ cm}^{-3}$. The carrier concentration in the p-region has a second-order effect compared with that of the n-region for the range of concentrations normally used.

Photocurrent and Quantum Efficiency

HgCdTe has a strong optical absorption coefficient and only thin layers are needed to produce high quantum efficiency. Typically in MW detectors the absorber need only be 4–5 μm thick and about twice this in LW detectors. Ideally the absorption should occur well within a diffusion length of the p–n junction to avoid signal loss due to recombination. A long carrier lifetime is nearly always observed in n-type material with low carrier concentration. Device engineers tend to favor using n-type absorbers for the best quantum efficiency and try to minimize the volume of the p-region for lower thermal leakage currents.

Excess Noise Sources in HgCdTe Diodes

There are many potential sources of excess noise in infrared detectors, and manufacturers strive to optimize obvious areas, such as the surface passivation (to limit surface leakage currents) and contacts. There are two other sources that have been reported in depth. The first

is linked to tunneling currents through traps, and the second is associated with crystal defects.

Many authors have reported an empirical relationship between tunneling leakage current and $1/f$ noise. The scatter within databases is usually large and the noise depends strongly on the technology used, but the noise trend is roughly given as αI_L^β , where α and β are variables depending on the device and leakage current mechanism. Leakage current from the trap-assisted tunneling mechanism results in β values close to 0.5, with α taking a value of $1 \times 10^6 \text{ A}^{0.5}$ [36.60]. With band-to-band processes the value of β moves towards 1. The nature of the trap has not been identified yet and it is not possible to exclude a variety of crystallographic defects. The physics of crystallographic defects in HgCdTe is complex because dislocations distort the local band structure and, via strain fields, the local bandgap, and this probably accounts for the variability in data in the literature.

The electrically active nature of dislocations and other crystal defects in HgCdTe is well reported [36.61, 62]. Dislocations can appear as n-type pipes and are associated with active defect centers. Many workers [36.63–68] have found that the reverse bias characteristics and $1/f$ noise of HgCdTe diodes depend strongly on the density of dislocations intercepting the junction. For HgCdTe this is an important observation because dislocations can easily be introduced during materials growth and the device fabrication process. Dislocations could increase the g–r current linearly in p-on-n heterostructures, along with a corresponding increase in $1/f$ noise current density [36.66]. Tunnel currents are also strongly associated with crystal defects, particularly at low temperatures.

The nature of defects in HgCdTe LW arrays has been studied in detail [36.58] in homojunction via-hole arrays (loophole arrays) made using high-quality LPE material, and a model has been proposed for the noise-generating mechanism. Threading dislocations that originate from the CdZnTe substrate and rise vertically through the layer can cause strong leakage in reverse bias if they intercept the junction, possibly because these dislocations can become randomly decorated during growth. Consequently, the threading dislocation density in the substrate is very important for controlling defect levels. Process-induced dislocations have a weaker effect on the junction properties resulting in an effective shunt resistance of about 40 M Ω , so an accumulation of this type of dislocation can result in a defect.

The current trend is to move towards lower temperature growth processes and growth on substrates with

poor lattice match, and this has refocused attention on grown-in dislocations. Many of these processes are not yet suitable for LW arrays. More data is needed for the VPE processes because the nature of the misfit dislocations and the geometry of the absorber will influence the magnitude of the excess noise.

36.6.4 Manufacturing Technology for HgCdTe Arrays

Considerable progress has been made over the last two decades in the epitaxial growth of HgCdTe. Bulk growth methods are still used to provide good quality material for photoconductor arrays, but for photovoltaic arrays there are problems associated with crystal grain boundaries, which are electrically active, and cause lines of defects. Also, there are limitations in the boule size, which makes it suitable for small arrays only.

Several epitaxial growth techniques are in use today. Manufacturers will select a technique that suits their device technology and the type of detectors they are trying to make. For instance, high-performance LW arrays will call for the best possible crystal quality, whereas large-area MW arrays can probably accept poorer material but must have large, uniform wafers. It is the aim of most manufacturers to produce high-quality layers in large areas at low cost, but this ideal has been elusive. At the present time the best structural quality material is grown using liquid phase epitaxy, LPE, onto lattice-matched crystals of CdZnTe, and this has been used successfully in homojunction technologies where the photosensitive junction is diffused into a homogenous monolith of material. There is a trend to move away from expensive CdZnTe substrates and both LPE and vapor-phase epitaxy (VPE) are now used on a variety of alternative substrates. Many groups favor VPE because the composition and doping profile can be easily controlled to produce complex devices, such as two-color detectors. The main HgCdTe growth processes are described here.

Summary of Growth

Using Liquid-Phase Epitaxy (LPE)

Liquid-phase epitaxy (LPE) of HgCdTe at present provides the lowest crystal defect levels, and very good short and long-range uniformity. LPE layers are grown using an isothermal supersaturation or programmed cooling technique or some combination. A detailed knowledge of the solid-liquid-vapor phase relation is essential to control the growth particularly in view of the high Hg

pressure. Challenges include: the compositional uniformity through the layer, the surface morphology, the incorporation of dopants and the specifications for thickness, wavelength, etc. A common component leading to high structural quality is the use of lattice-matched substrates of CdZnTe. These are grown by a horizontal Bridgman process and can supply layers as large as 6×4 cm. The CdZnTe substrates must be of the highest quality, and often this is a significant cost driver for the process.

Two different technical approaches are used: growth from a Hg-rich solution, and growth from a Te-rich solution. Advantages of the Hg-rich route include: excellent surface morphology, a low liquidus temperature, which makes cap layer growth more feasible, and the ease of incorporation of dopants. Also, large melts can provide for very good compositional and thickness uniformity in large layers and give consistent growth characteristics over a long period of time. Growth from Te-rich solutions can use three techniques to wipe the melt onto the substrate: dipping, tipping and sliding. Sliding boat uses small melt volumes and is very flexible for changing composition, thickness and doping. Tipping and dipping can be scaled up easily and can provide thick, uniform layers but the large melts limit flexibility. Double layers are also more difficult to grow.

Most manufacturers have taken their chosen growth system and tailored it to provide optimum material for their device technology. In particular the use of dopants and the deliberate introduction of compositional grades are very specific to the device structure. A crucial figure of merit however is the dislocation count, that controls the number of defects in 2-D arrays. Etch pit densities of $3\text{--}7 \times 10^4 \text{ cm}^{-2}$ are typically seen in the Te-rich sliding boat process reproducing, the substructure of the CdZnTe substrate [36.69]. The etch pits are associated with threading dislocations which appear to be normal or near normal to the layer surface. The substrate defect level can be as low as $\text{mid-}10^3 \text{ cm}^{-2}$ in some horizontal Bridgman CdZnTe, but this is not easy to reproduce. Similar defect densities are found in CdTeSe material but the impurity levels have proved difficult to control in the past. Device processing therefore must expect to cope with defect levels in the $\text{mid-}10^4 \text{ cm}^{-2}$ range for routine CdHgTe epilayers, and this will set the ultimate limit on the number of defects in HgCdTe 2-D arrays. Several groups have used the LPE process on low-cost substrates, including: CdZnTe or CdTe on GaAs/Si wafers [36.70] or the PACE technology on sapphire described in detail in Sect. 36.6.4.

Summary of Growth Using Metalorganic Vapor-Phase Epitaxy (MOVPE)

A detailed summary of the state of the art for MOVPE technology has been produced [36.71], but this section summarizes the main points.

MOVPE growth depends on transporting the elements Cd and Te (and dopants In and As) at room temperature as volatile organometallics. They react along with Hg vapor in the hot gas stream above the substrate or catalytically on the substrate surface. The drive to lower temperatures and hence lower Hg equilibrium pressures has resulted in the adoption of the Te precursor di-isopropyl telluride, which is used for growth in the 350–400 °C range. A key step in the success of this process is to separate the CdTe and HgTe growth so that they can be independently optimized. This is called the IMP process (interdiffused multilayer process) [36.72]. IMP results in a stack of alternating CdTe and HgTe layers and relies on the fast interdiffusion coefficients in the pseudobinary to homogenize the structure at the growth temperature. Doping is straightforward using Group III metals (acceptors) and Group VII halogens (donors). For instance, ethyl iodide is used for iodine doping.

The main morphological problem for MOVPE are macro defects called hillocks, which are caused by preferred 111 growth, nucleated from a particle or polishing defect. Hillocks can cause clusters of defects in arrays. Orientations 3–4° off 100 are used primarily to reduce both the size and density of hillocks.

A variety of device structures with layers have been reported [36.73] based on MOVPE-grown HgCdTe layers on 75 mm-diameter GaAs on silicon.

Summary of Growth Using Molecular Beam Epitaxy (MBE)

A good detailed summary of the state of the art for MBE technology has been reported [36.74].

MBE offers the lowest temperature growth under an ultrahigh vacuum environment, and, in common with MOVPE, in situ doping and control of the composition and interfacial profiles. These are essential for the growth of advanced and novel device structures. Typically growth is carried out at 180–190 °C on 211 CdZnTe substrates. Effusion cells of CdTe, Te and Hg are commonly used. Hg is incorporated into the film only by reacting with free Te, and so the composition depends on the Te to CdTe flux ratio. The structural perfection depends strongly on the Hg to Te flux ratio and growth is usually restricted to a tight temperature range. Indium is the most widely used n-dopant and is well activated. p-type dopants are less conveniently incorpo-

rated in situ but manufacturers have devised a number of processes to force As onto the proper Te site. Again the Hg to Te ratio and growth temperature is crucial to achieving good activation. In general, reproducibility seems more difficult to achieve than MOVPE. MBE structural problems center mainly on pinholes or voids. Some very good EPD levels have been reported, but in general the EPD levels are an order or more higher than the best quality LPE.

Junction Forming Techniques

For n–p devices, crystal growers can obtain the desired p-type level by controlling the density of acceptor-like mercury vacancies within a carrier concentration range of, say, 10^{16} to 10^{17} cm⁻³. Neutralizing the Hg vacancies and relying on a background level of donors to give the n-type conversion creates the photodiode junctions. Mercury can be introduced by thermal diffusion from a variety of sources, but high-temperature processes are not very compatible with HgCdTe at the device level. However, type-conversion can be readily achieved by processes such as ion beam milling [36.75] and ion implantation [36.76–81]. Type conversion can also be achieved using plasma-enhanced milling in the VIAP process [36.81] and also using H₂/CH₄ plasmas [36.82]. The common feature is that the conversion depth is much deeper than would be expected from the implantation range alone.

The current knowledge on type conversion using ion beam milling is described by [36.58] and the current knowledge for ion implantation is summarized by [36.83]. The explanation for the behavior of HgCdTe under ion beam bombardment involves a number of physical mechanisms. Firstly, the low binding energies, ionic bond nature and open lattice of HgCdTe encourages the liberation of free mercury at the surface and subsequent injection by the ion beam. The injection mechanism probably involves a recoil implantation process. Once the Hg interstitial is injected, the mobility is apparently extremely high and there is some evidence that this is stimulated by the ion beam in a process related to the anomalously high diffusion rates of impurities often observed in SIMS analysis. Another factor is the movement of dislocations under the influence of the ion beam and possibly stimulated by strain fields. A number of workers report that the n-type carrier concentration in the converted region is very low, and in fact fast-diffusing impurities such as the Group IB elements Cu, Ag and Au and the Group IA elements Na, K and Li, which reside on the metal sublattice, are swept out of the n⁻ region of the diode by the flux of Hg interstitials. This impurity

sweep-out effect is serendipitous because it means that in ion beam generated junctions the n-regions are very pure, free of S-R centers and have weak doping, creating an ideal structure for high-performance detectors. The detailed atomic level processes taking place including the role of mercury interstitials, dislocations and ion bombardment in the junction-forming process are complex and not well-understood in detail. Despite the complex physics involved, manufacturers have achieved good phenomenological control of the junction depth and n-dopant profiles with a variety of processes.

Via-Hole Technologies

Via-hole devices share many of the process stages of photoconductive arrays and have been commercially produced since 1980. Arrays made using this process are known as loophole devices [36.84] and HDVIPTM devices [36.81]. By using a thin monolith of HgCdTe bonded rigidly to the silicon, the thermal expansion mismatch problem is overcome because the strain is taken up elastically. This makes the devices mechanically and electrically very robust and suitable for very long linear arrays, such as a 512×2 array of over 16 mm in length [36.69].

The process flow of the via-hole device using LPE material is illustrated in Fig. 36.12 [36.59, 69]. The process has two simple masking stages. The first defines a photoresist film with a matrix of holes of, say, 5 μm

diameter. Using ion beam milling, the HgCdTe is eroded away in the holes until the contact pads are exposed. The holes are then backfilled with a conductor, to form the bridge between the walls of the hole and the underlying metal pad. The junction is formed around the hole during the ion beam milling process, as described in Sect. 36.6.4. The second masking stage enables the p-side contact to be applied around the array. Figure 36.12 shows a photomicrograph of one corner of the hybrid, illustrating the membrane-like nature of the HgCdTe on the silicon. The device structure inherently produces low crosstalk in small pixels.

The HDVIPTM process [36.81] results in a similar structure but uses a plasma etching stage to cut the via-hole and an ion implantation stage to create a stable junction and damage region near the contact. In order to achieve higher lifetimes and lower thermal currents, Cu is introduced at the LPE growth stage. This is swept out during the diode formation and resides selectively in the p-region, partially neutralizing the S-R centers associated with Hg vacancies and achieving dark currents approaching those of fully doped heterostructures.

In via-hole processes the cylindrical shape of the junction minimizes the intersection with threading dislocations (Sect. 36.6.3). This has been demonstrated by the difference in yields when the HDVIPTM device is extended by a planar junction [36.81].

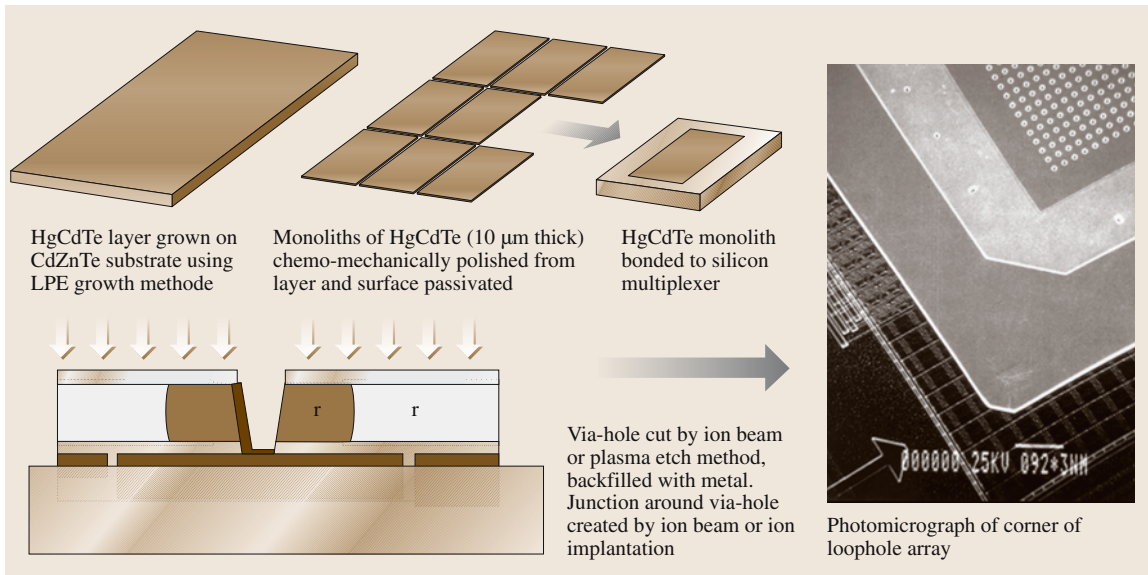


Fig. 36.12 Processing of HgCdTe arrays using LPE material and via-hole technology

Planar Device Structures Using LPE

The planar device structure illustrated in Fig. 36.13 is the simplest device structure currently used. It is consistent with a number of junction forming processes, e.g. ion implantation, diffusion and ion milling. The matrix of junctions is mass connected to an underlying silicon multiplexer using indium bumps. The strength of the process is the simplicity and the compatibility with epitaxially grown materials. A process based on high-quality LPE material and ion-implanted junctions is conducive to good-quality detectors and volume production [36.85]. In the simplest form, the process needs three masking stages for the junction and pixel contact, and one for the contact to the p-side. The device is backside-illuminated, i.e. it is illuminated through the substrate, and so this must be of high optical quality. Careful control of the junction geometry is needed to avoid crosstalk due to the diffusion of minority carriers into adjacent pixels, especially in the case of small pixel sizes. The thermal expansion mismatch between the HgCdTe/CdZnTe substrate combination and the silicon multiplexer is another important consideration in this device structure and this can restrict the practical size of the array unless the CdZnTe substrate is thinned.

Double Layer Heterojunction Devices (DLHJ)

Thermal leakage currents in HgCdTe devices tend to arise from the p-type side, and most advanced technologies strive to improve the operating temperature by using the n-type side as the absorber. n-type HgCdTe is easier to control at low carrier concentrations and is relatively free of Schottky-Read centers that limit the lifetime in p-type material. The p-type side then is often minimized

in volume or uses a wider bandgap to reduce the thermal generation. For heterostructures that are grown in situ, it is necessary to completely isolate the junction, forming a so-called mesa device. This is usually performed using a chemical etch because of the electrical side effects of using dry processing techniques.

When a device uses different compositions or doping levels, a nomenclature is adopted to describe the structure. The definition used in this chapter is as follows. An n-type diode formed in a p-type layer is described as an n-p device, so the first letter is the layer nearest the contact and photons are absorbed in the p-type layer. Wider bandgap material is indicated by using a capital letter. A superscript “-” or “+” denotes particularly low or high doping.

Double-layer heterojunction devices have been developed mainly in the US for long-wavelength detectors with low thermal leakage currents (or high R_0A values). A number of elegant device structures have been reported with R_0A values that are an order of magnitude higher than those of via-hole or planar diodes. The back-illuminated mesa P^+-n heterojunction, illustrated in Fig. 36.14, is a widely used device, and has been reported from both LPE and MBE material. This structure makes good use of the n-type absorber to give a high fill factor. Devices have been made using a vertical dipper LPE process from a Hg rich solution [36.86] and using a horizontal slider LPE process from a Te-rich solution [36.87–89]. The p-type layer is doped with arsenic to around $1-4 \times 10^{17} \text{ cm}^{-3}$ and is grown by the vertical dipper LPE process from a Hg-rich solution. An n^-P^+ structure has been reportedly grown from MBE which has the advantage of very low crosstalk because of the complete electrical isolation of the absorber [36.90].

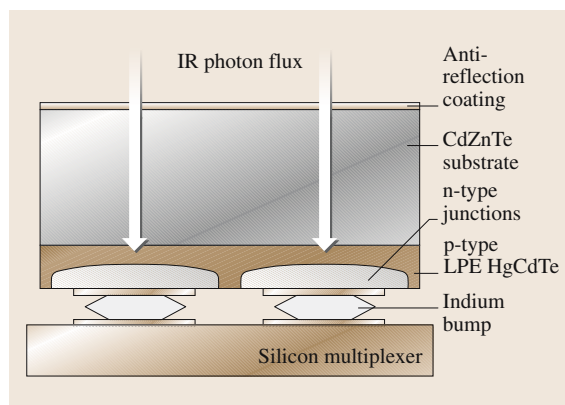


Fig. 36.13 Schematic of planar indium-bumped hybrid

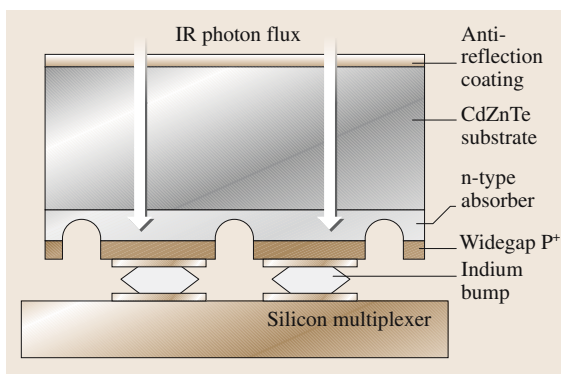


Fig. 36.14 Schematic of double-layer heterojunction (DLHJ)

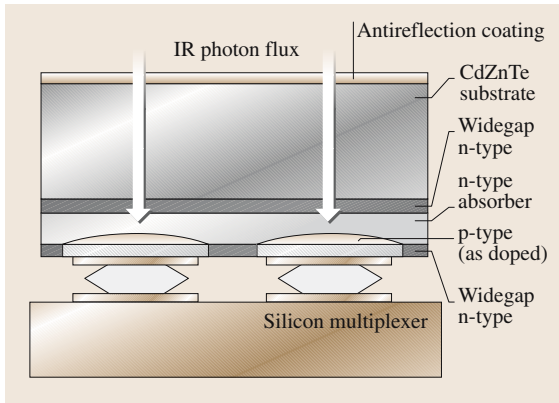


Fig. 36.15 Schematic of double-layer planar heterostructure (DLPH)

Other workers have produced the conventional P^+n^- structure using an MBE process [36.91].

Mesa heterostructures have a technologically difficult passivation stage on the sidewalls of the mesa, which can lead to reverse bias leakage and uniformity problems if not optimum. In order to solve the sidewall problem, a clever diffused heterostructure has been reported [36.92], called a double-layer planar heterostructure (DLPH). This is a back-illuminated arsenic-implanted p-n-N planar buried junction structure [36.93], as illustrated in Fig. 36.15. The junction is formed by arsenic implantation into a three-layer N-n-N film grown in situ by MBE onto a CdZnTe substrate. The unique feature of this structure is that the junction is buried beneath the top wide-bandgap n-type layer. The junction intercepts the surface in wider gap CMT, thereby reducing the generation rates for any surface defects that may be present. The arsenic doped region must extend deep enough into the narrow gap absorber layer in order to collect photocarriers. This necessitates that the wide gap layer be thin (on the order of $0.5\ \mu\text{m}$) and that the compositional interdiffusion between the wide gap and absorber be minimal. Growth by MBE at low temperature (175°C) satisfies these requirements. Another feature of this structure is the wide-gap n-type buffer layer between the absorber and the substrate. This buffer layer keeps carriers that are photogenerated in the absorber away from the film-substrate interface where there can be interface recombination. The DLPH process has been used to produce arrays of up to 1024×1024 on $18.5\ \mu\text{m}$ pitch (HAWAII multiplexers) with high quantum efficiency [36.94].

Wafer-Scale Processes Using Vapor Phase Epitaxy on Low-Cost Substrates

There is a dual purpose in developing vapor phase epitaxy technologies. Firstly, it opens up the possibility of growing complex multilayer materials for a whole range of new device structures, and secondly it allows growth on lower cost substrates. Certainly the elegance of this technique is that the composition, thicknesses and doping levels can be programmed, in principle, to grow any required detector structure. For instance, multilayer, fully doped heterostructures have been grown which demonstrate low thermal leakage current by a process called Auger exclusion [36.95]. Also, two-color detector structures can be grown as described in Sect. 36.6.8. The main commercial drive though is to cost-reduce the manufacture of large-area arrays and avoid expensive CdZnTe substrates. Silicon, gallium arsenide and sapphire are the commonest substrates used because they are available commercially in large-area wafers at relatively low cost. Figure 36.16 illustrates a typical process flow.

One of the main considerations when designing large arrays is the thermal mismatch problem between the various layers in the device and carrier, which can cause

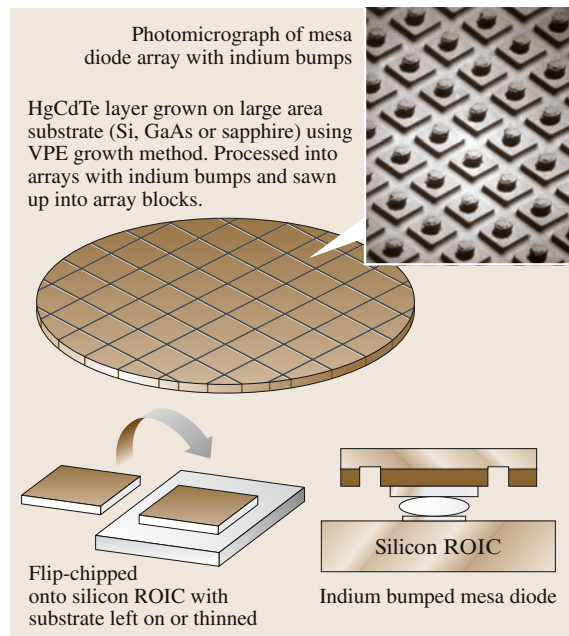


Fig. 36.16 Wafer-scale processing using VPE HgCdTe material, mesa diodes and indium bump interconnect technology

bowing or cracking at low temperature. The silicon multiplexer has a very low thermal expansion coefficient compared to most materials used and a number of different techniques are used to cope with this.

Some groups focus on the growth of HgCdTe on silicon substrates so that, when bump-bonded, the substrate and multiplexer are balanced. Silicon is also mechanically strong, available in large wafers at low cost, and avoids the recurring problem of impurity out-diffusion in CdZnTe substrates. A 75 mm-diameter silicon wafer can produce about twelve 640×512 arrays. For HgCdTe there is a 19% lattice constant mismatch with silicon, and this can potentially provide high levels of misfit dislocations, resulting in degraded performance, particularly at longer cut-off wavelengths. However at short and medium wavelengths the effect of misfit dislocations is reduced enough to make these wafer-scale processes viable at the present time. Most of the MOVPE work on silicon has used a GaAs layer to buffer the lattice mismatch between silicon and HgCdTe [36.73], and a variety of device structures are reported using MOVPE layers grown on 75 mm-diameter GaAs on silicon layers. MBE growth on silicon substrates of 100 mm in diameter has been used to grow MW HgCdTe layers [36.96]. The MBE work has used buffer layers of CdZnTe and CdTe to minimize the density of misfit dislocations. Both MBE and MOVPE growth has been used to produce LW arrays on misorientated silicon substrates [36.97]. An MBE CdTe/ZnTe layer is first grown and this acts as a buffer for the HgCdTe growth using an MOVPE process.

Another approach, known as PACE, has been described [36.98]. The process uses a 75 mm-diameter substrate of highly polished sapphire, which remains

with the device throughout. The thermal expansion problem is overcome by using a carrier substrate of sapphire so that the silicon multiplexer is sandwiched in the middle. The process starts with a buffer layer of CdTe deposited using an organometallic vapor-phase epitaxy (OMVPE) process to $8.5 \mu\text{m}$ thickness followed by a $10 \mu\text{m}$ -thick p-type HgCdTe layer grown by LPE. The n-p diodes are formed by ion implantation. This process has been used to fabricate 1024×1024 arrays with $3.2 \mu\text{m}$ cut-off [36.42] and 2048×2048 SW arrays of 16 cm^2 in area [36.99]. The 2048×2048 array, produced by Rockwell Scientific Company LLC for an astronomy application, is shown in Fig. 36.17. The arrays are $37 \times 37 \text{ mm}$ in size and are backside-illuminated through the sapphire substrate.

36.6.5 HgCdTe 2-D Arrays for the 3–5 μm (MW) Band

In the MW band, the development of array technology has reached a stage where all of the reported device processes will result in essentially background-limited performance for a wide range of applications. Individual arrays differ from one another in array size and pixel size and in imaging artefacts, such as uniformity, crosstalk and blooming. Some mention needs to be made of the factors controlling sensitivity in 2-D arrays, because this does not depend primarily on the HgCdTe but the silicon multiplexer.

Both charge-coupled devices (CCDs) and standard foundry CMOS devices have been used for multiplexing infrared 2-D arrays, but CMOS is now the preferred choice and CMOS circuits will operate well at low temperatures. The function of the multiplexer is to integrate the photocurrent from the photodiode array, perform a limited amount of signal processing and sequentially readout the signals. Most manufacturers produce their own multiplexer designs because these often have to be tailored to the application. For background-limited detectors, the NETD (noise-equivalent temperature difference) can be shown to be inversely proportional to the square root of the number of integrated photogenerated electrons [36.100]. High sensitivity can therefore only be achieved if a large number of electrons are integrated and this requires the integration capacitance in each pixel to be fairly high. Normally the capacitor is created using a MOSFET structure with a thin gate oxide dielectric and capacitance densities as high as $3 \text{ fF } \mu\text{m}^2$ are typical. Nevertheless, the capacitance is restricted to about 1 pF in pixels of around $25 \mu\text{m}$ square, and so the best NETD that can be expected is about 10 mK per

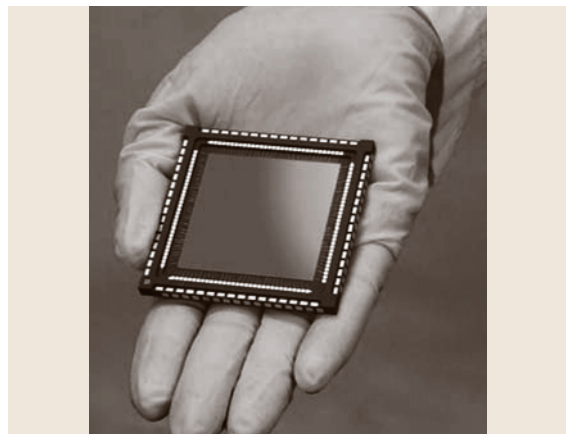


Fig. 36.17 2048×2048 SW array for astronomy. (Courtesy of Rockwell Scientific Company)

frame. It is important to note that this NETD limitation has nothing to do with the HgCdTe but only depends on the multiplexer design. Furthermore, as the pixel size is reduced the capacitance is decreased disproportionately, so the NETD cannot be maintained in small-pitch arrays.

Some general comments about operating temperature can be made. Standard vacancy-doped HgCdTe diodes, such as planar or via-hole homojunctions, need cooling to 120–140 K for BLIP performance in typical thermal imaging applications. The standard HDVIPTM, which is extrinsically doped with many fewer vacancies, can operate up to 155 K. HgCdTe diodes fabricated with controlled doping and bandgap engineering can achieve BLIP operation near to ≈ 200 K. Operating at a higher temperature can prolong the life of cryocoolers and results in lower power consumption.

HgCdTe offers better spatial resolution (often quantified as the modulation transfer function, MTF) than some competing technologies such as planar InSb processes. Fully defined MOVPE or MBE diode structures will have the best MTF performance, but via-hole technologies (loophole and HDVIP) closely approach them because of their concentric junction shape, and this permits true resolution improvement in small pixels. DLHJ structures and planar structures rely on lateral collection to give very high fill factor, and high quantum efficiencies, and they need to have restricted sideways diffusion to limit blurring. Blooming due to an intense optical highlight follows a similar argument.

36.6.6 HgCdTe 2-D Arrays for the 8–12 μm (LW) Band

LW HgCdTe detectors were originally developed for long linear arrays and the processes can be extended to 2-D arrays with one additional consideration. The earth plane of the array must carry the sum of the photocurrent from all of the pixels in the array, and if there is insufficient conductance a voltage will appear between the center and edge of the array. This is called “substrate debiasing”, and in the extreme case will inhibit injection in the center diodes. In any event it provides a potential long-range crosstalk mechanism and detector designers will try to minimize the resistance of the earth plane layer. In homojunction arrays, the p-type common layer cannot be doped high enough and a special metal grid is needed, as shown in the DRS Technologies array in Fig. 36.18. In other technologies the common layer must be doped high enough to prevent long-range crosstalk and use junctions with high dynamic resistance.

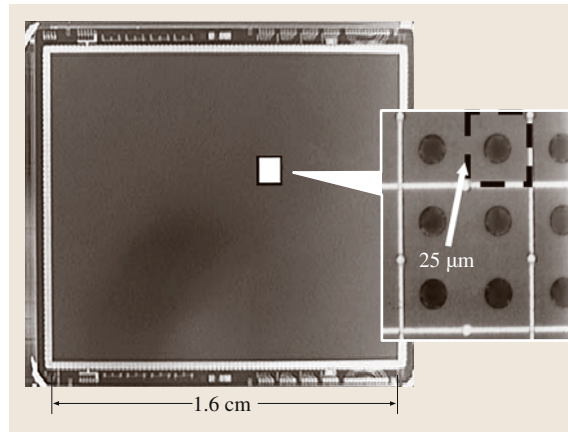


Fig. 36.18 LW HDVIP array with metal grid for substrate debiasing control. (Courtesy DRS Technologies)

36.6.7 HgCdTe 2-D Arrays for the 1–3 μm (SW) Band

Important applications in the SW (1–3 μm) wave-band include: thermal imaging (using nightglow), spectroscopy and active imaging using lasers. Other materials can be used for SW detectors but products such as InGaAs tend to have high noise beyond 1.7 μm due to defects arising from the lattice mismatch with the InP substrate. Most HgCdTe technologies can be extended to SW with little change to the processing. Carrier lifetimes should be dominated by radiative recombination, but in fact S-R centers and technologically related limits probably apply in practice. SW detectors can produce good quality diodes in the presence of fairly high levels of misfit dislocations and so can be made using some of the newer technologies described in Sect. 36.6.4. For imaging and spectroscopy applications, typical operating temperatures are around 200 K and thermoelectric coolers are often offered as standard products.

A growing area is active imaging using eyesafe 1.54 μm lasers. The shorter wavelength results in higher resolution for a given optical aperture size and is particularly beneficial in long-range sensor systems. The very short pulse nature of the laser can be used to give range information and is often used with gated detectors to produce imaging in a certain range zone (so-called laser shape profiling). When range information is produced it is called laser range profiling and, in the limit, 3-D imaging. At present the field is in its early development stage and workers are using relatively small arrays to perform initial research. Detectors for this field need to be very fast and use a wide, weakly doped absorber

region to give low capacitance. This is a so-called PiN diode, where the “i” stands for intrinsic. It also needs a very low series resistance to withstand the shock of the laser return, and this lends itself to metal grids and heavily doped heterostructures.

In most SW applications the photon flux is low and it is difficult to achieve reasonable signal-to-noise performance. However, it is relatively easy to enhance the signal by providing some avalanche gain in the device. Electron avalanching in MW HgCdTe via-hole diodes has been described [36.101, 102]. The electron and hole ionization rates are very different in HgCdTe, and this allows almost pure exponential, noise-free avalanche gain at fairly moderate voltages. So avalanche gain can readily be achieved for wavelengths above about $2.5\text{ }\mu\text{m}$ when the absorber region is p-type and electrons are the minority carriers. A gain of 10 is typically observed for around 5 V at a cut-off of $5\text{ }\mu\text{m}$. An alternative structure uses SW material ($1.6\text{ }\mu\text{m}$) and a resonant enhancement of the hole impact ionization rate when the bandgap equals the spin-orbit split-off energy [36.103]. Gains of 30–40 have been seen with voltages of 85–90 V.

SW laser-gated imaging systems using avalanche gain in HgCdTe are now being reported for use in long-range identification applications [36.104]. Here the combination of sophisticated ROICs and high quality HgCdTe device processing is producing 320×256 arrays with a sensitivity down to 10 photons rms.

36.6.8 Towards “GEN III Detectors”

“GEN III” is a commonly used term which stems from the abbreviation for third-generation infrared detector. The definition of a “GEN III” detector can differ between different nations but the general guideline is any detector that offers an imaging advantage over conventional first- and second-generation systems. Common agreed examples include megapixel arrays with high density, dual color or even multispectral arrays, higher operating temperature, fast readout rates, very low NETD due to pixel-level signal compression and retina-level signal processing. Some of the progress made in such research is described here in more detail.

Two-Color Array Technology

Resolving the spectral signature can enhance the identification of objects in a thermal image. As a stepping stone to true multispectral arrays, many manufacturers are developing two-color detectors with simultaneous readout of flux levels in two separate infrared bands. A good example of a field requirement is to separate

a sun glint from a hot thermally emitting object. The field is still in its early development stages, but some good results have been achieved.

Two-color arrays use two layers of HgCdTe with the longer wavelength layer underneath. The technological challenge is to make contact with the top layer without obstructing the sensitive area of the bottom layer. Ideal devices, where the integration takes place simultaneously and the sensitive area is co-spatial, are practically difficult to make on normal pixel sizes of say less than $50\text{ }\mu\text{m}$. Most dual-color devices using indium bumps need two contacts within each pixel and make contact to the top diode by etching a hole in the bottom layer. The silicon needs to be custom-designed because the flux levels in the two bands may be markedly different. The polarity of the input MOSFET and the gain within the silicon must be matched to the technology and application.

Several successful arrays have been reported including a 64×64 simultaneous MW/LW dual-band HgCdTe array on $75\text{ }\mu\text{m}$ pitch [36.105]. The array is fabricated from a four-layer P-n-N-P film grown in situ by MOVPE on CdZnTe and mounted on a custom ROIC. This approach has been applied to the double-layer planar heterostructure, DLPH, in a process called SUMIT, which stands for simultaneous unipolar multispectral integrated technology [36.106]. The DLPH structure can be turned into a two-color device by employing two absorbers with different bandgaps separated by a wide bandgap barrier. An etch step is used to expose the lower, shorter wavelength absorber and to form an arsenic doped junction. A pixel size of $40\text{ }\mu\text{m}$ for each color is used in 128×128 demonstrators for MW/MW arrays. A variant of this basic process has also been presented by AIM [36.107] and they have achieved two colors in a concentric arrangement in 192×192 and 256×256 arrays with $56\text{ }\mu\text{m}$ pitch, again using a full custom ROIC.

Two-color arrays have been fabricated using a via-hole type structure in which each pixel has two via-holes [36.81]. One is for connecting the longer wavelength array to the silicon in the normal way and the other is isolated from the longer wavelength material and connects the shorter wavelength to the silicon, as is illustrated schematically in Fig. 36.19. The spectral response of the two colors is shown in Fig. 36.20 for the two different doublets. The two-color arrays utilize standard 640×480 , $25\text{ }\mu\text{m}$ pitch ROICs (as used in LW focal plane arrays). The unit cell design in these ROICs incorporates one buffer amplifier circuit for every two HgCdTe pixels. The two-color array actually uses four

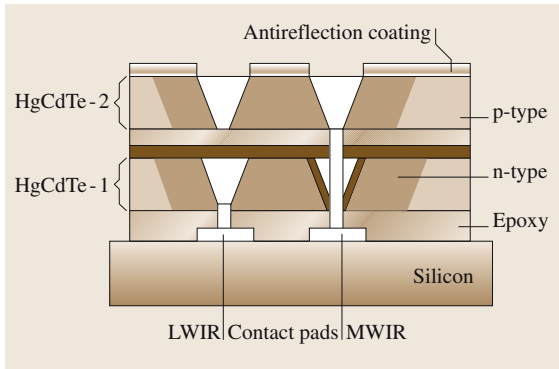


Fig. 36.19 MW/LW HDVIP FPA two-color composite. (Courtesy DRS Technologies)

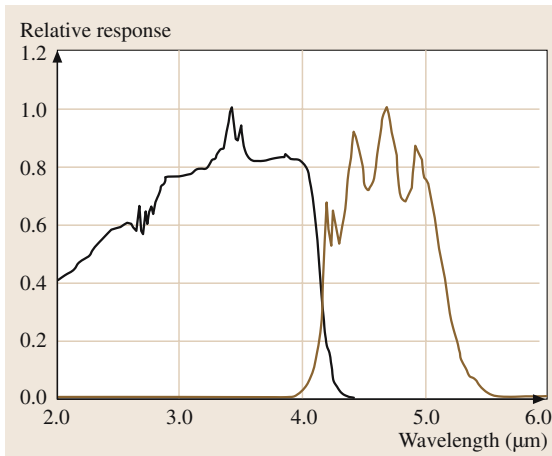


Fig. 36.20 Spectral response of MW/MW two-color HDVIP HgCdTe FPA. (Courtesy DRS Technologies)

mono-color pixels to form a two-color unit cell, thus resulting in a $50\text{ }\mu\text{m} \times 50\text{ }\mu\text{m}$ geometry, with two buffer amplifiers per unit cell, one for each color.

The shrink potential of dual-color technologies is limited by the need for two contacts per pixel. An alternative approach is to use a so-called bias-selectable detector [36.108–110], which needs only one bump and is compatible with small pixel sizes. This is an elegant technique but has the operational disadvantage of nonsimultaneous integration, and for this reason these detectors are often called sequential mode, dual-band detectors. Typical devices use a p-n-N-P or n-p-P-N sandwich structure with each p-n junction in different composition material. The wavelength can be selected by the bias polarity. There is a potential problem with parasitic transistor action and this needs to be suppressed

by appropriate barriers. The silicon circuit requires a PMOS input for one polarity and an NMOS for the other and these must be switched for each color.

Higher Operating Temperature (HOT) Device Structures

The general principles for designing HOT HgCdTe detectors have been proposed over the past two decades, primarily as a result of research in the UK. Good summaries have been reported in [36.111, 112]. The advantages afforded by HOT detectors are in the area of cost of ownership and portability. LW arrays operating at only a few tens of degrees higher can significantly improve the mean time before failure (mtbf) of a Stirling engine cooler. In the MW band BLIP detectors on thermoelectric coolers could offer a much cheaper, smaller and more robust product, ideally suited to continuously operating security-type applications. Operation at 220 K has been demonstrated for a 128×128 array using MOVPE heterodiode technology [36.113], and the key step was designing a multiplexer that enabled very accurate biasing of the diodes at zero volts to suppress the dark current and associated $1/f$ noise.

The possibility of operating detectors at temperatures near room temperature has been proposed [36.114]. The basic structure is a P-p⁻N, as illustrated in Fig. 36.21. The active volume of the absorber is small relative to a minority carrier diffusion length, and it is operated in strong nonequilibrium by reverse biasing the minority carrier contact to completely extract all of the intrinsically generated minority carriers. To preserve space-charge neutrality, the majority carrier concentration drops to the background dopant concentration and Auger generation is effectively suppressed. Remaining components are associated with S-R centers and possibly injection from the contact and surface regions. The realization of HOT LW arrays will depend on the suppression of high $1/f$ noise associated with the reverse bias operation. In MW arrays it is also important to

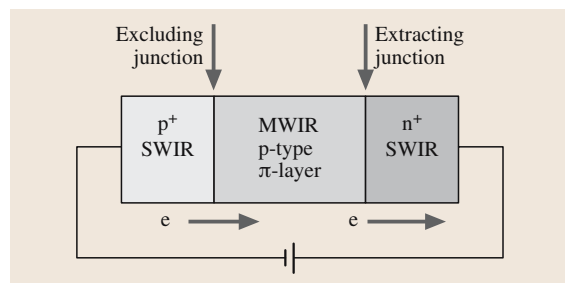


Fig. 36.21 Schematic of HOT device structure

suppress competing recombination mechanisms. Nevertheless, the device concept has enormous importance as a route towards high-performance infrared detectors with minimal cooling.

Retina-Level Processing

In very large arrays operating at high frame rates, the downloading of signal data and the subsequent signal processing can be daunting and is called the data processing bottleneck problem. The human retina presents an example of how evolution has dealt with the problem. The eye performs a number of image processing operations in the “z-plane”, including, in order: logarithmic photon sensing, spatial filtering, temporal filtering, motion sensitivity and data decomposition. An essential function is the Difference of Gaussian or DoG filter which is used for both edge and contrast enhancement [36.115].

Focal plane arrays with neuromorphic processing are under development at a number of centers [36.116–118]. A good example is the MIRIADS programme (miniature infrared imaging applications development system). MIRIADS uses a neuromorphic FPA, with temporal high-pass filtering, frame co-adding and a Difference of Gaussians operation to detect motion, enhance edges and reject ambient light levels. The current reported array is a 64×64 .

36.6.9 Conclusions and Future Trends for Photovoltaic HgCdTe Arrays

Photovoltaic 2-D HgCdTe detectors are serving many applications worldwide and are supplanting photoconductive detectors in second-generation systems for military, commercial and scientific applications, offering improved temperature sensitivity, lower power consumption, weight and volume. The continuing development work within manufacturing centers, supported

by the wider scientific community, is aimed at increasing the radiometric performance, reducing defects and reducing manufacturing costs. Medium-term research is being directed at the best advanced materials and device structures for the next generation of HgCdTe arrays, which will combine state of the art background limited performance with low-cost wafer-scale processing. In the future it is envisaged that HgCdTe detectors will be grown directly on silicon or even silicon multiplexers for very low cost detectors. Bandgap engineering will produce heterostructure detectors with much higher operating temperatures. This will enable background-limited operation in MW arrays at near room temperature and operation of LW detectors on thermoelectric coolers or low-power Stirling engines. Device structures will be extended to produce bispectral and multispectral capability. Advanced detectors will critically depend on the future development of multiplexers. Challenges for the future include: arrays with larger physical size, smaller pixels, higher sensitivity, faster frame rate and even perhaps retina-type processing to reduce the need for external signal processing. The ultimate performance potential of HgCdTe will ensure that it is the material of choice for all high-performance infrared systems. In MW arrays it will compete with InSb offering better imaging characteristics and higher operating temperature. Active imaging will drive a major growth in SW array technology. The top performing thermal imaging cameras currently depend on LW, HgCdTe, long linear arrays offering high-definition TV images: the only technology currently capable of doing this economically. The market for smaller LW 2-D arrays is dominated by uncooled (thermal) detectors and multiple quantum well detectors, but HgCdTe offers considerably improved sensitivity, and once the cost becomes competitive it should compete with these technologies for many applications.

36.7 Emission Devices in II–VI Semiconductors

Infrared emission devices, in the form of LEDs (light emitting diodes) and laser diodes, are used in a broad range of applications such as gas sensors, infrared scene simulators, molecular spectroscopy, free space communications, fiber optic communications and LIDAR (light detection and ranging). A useful summary of emission device theory and current progress has been presented by [36.119]. There is competition from lead salt semiconductors (in the 3 to $34\text{ }\mu\text{m}$ range) and narrow gap

III–V compounds. HgCdTe LEDs use a forward-biased junction to inject minority carriers into an n-type region, typically with low doping to encourage radiative recombination. The external photon efficiency is limited by factors such as competing recombination mechanisms, reabsorption of the photons and internal reflection. Nevertheless, efficiencies as high as 5% have been achieved at $4\text{ }\mu\text{m}$, using an immersion lens to overcome internal reflection [36.120].

One of the disadvantages of the LED is that the spectrum is often too broad for many applications, and lasers are needed. Lasers also offer greater output power and higher spatial coherence, but they are more difficult to make and have additional noise mechanisms. Infrared lasers covering the 3–5 μm band have been demonstrated with pulsed operation at 80 K. It is only recently that the complex double heterojunction structures needed for lasing have been available from low-temperature VPE processes, so the field is relatively young. Nevertheless, peak powers of up to several milliwatts have been recorded in diode pumped lasers and over 1 W in optically pumped lasers.

Using a structure similar to that reported in Sect. 36.6.8 and illustrated in Fig. 36.21, the carrier densities, in reverse bias, can be lowered to well below the thermal equilibrium values. Under this condition, the absorption of photons can exceed the emission rate and the surface of the device can look cold. This is called negative luminescence [36.121] and devices have been developed for use as temperature references in thermal imaging cameras.

Infrared LEDs have also been reported [36.122] in the dilute magnetic semiconductors $\text{Hg}_{1-x}\text{Mn}_x\text{Te}$ and $\text{Hg}_{1-x-y}\text{Cd}_x\text{Mn}_y\text{Te}$, where, in principle, the bandgap could be tuned by the applied magnetic field.

36.8 Potential for Reduced-Dimensionality HgTe–CdTe

The emergence of low-temperature growth processes will enable HgTe–CdTe structures to be grown with dimensions small enough to show quantum confinement effects. At present this research is in its early stages, and there are no active programs to make devices, but this could change in future. HgTe–CdTe superlattices offer many benefits for detectors. The minority carrier lifetime could be enhanced beyond that of bulk material, providing better quantum efficiency. The optical absorption coefficient can be much larger than bulk material, allowing LW detectors to use absorbers as thin as

a few microns. Furthermore, tunneling currents can be reduced by orders of magnitude, permitting operation at much longer wavelengths. The state of the art of reduced dimensionality material is summarized in [36.123].

Quantum well lasers are also known to offer significant advantages over double heterojunction devices with bulk active regions. For instance, the maximum operating temperatures for pulsed and CW lasers could be increased by over 100 K, and this would have important consequences for the attractiveness of narrow-gap II–VI emitters.

References

- 36.1 W. D. Lawson, S. Nielsen, E. H. Putley, Y. S. Young: J. Phys. Chem. Solids **9**, 325–329 (1959)
- 36.2 M. A. Kinch: Mater. Res. Soc. Symp. Proc. **90**, 15 (1987)
- 36.3 N. Duy, D. Lorans: Semicond. Sci. Technol. **6**(12), C93 (1991)
- 36.4 S. Oguz, R. J. Olson, D. L. Lee et al.: Proc. SPIE **1307**, 560 (1990)
- 36.5 T. Tanaka, K. Ozaki, K. Yamamoto et al.: J. Cryst. Growth **117**, 24 (1992)
- 36.6 M. A. Kinch, S. R. Borello: Infrared Phys. **15**, 111 (1975)
- 36.7 C. T. Elliott: *Handbook on Semiconductors*, 1st edn., ed. by C. Hilsum (North-Holland, Amsterdam 1981) p. 727
- 36.8 C. T. Elliott, N. T. Gordon: *Handbook on Semiconductors*, 2nd edn., ed. by C. Hilsum (North-Holland, Amsterdam 1993) p. 841
- 36.9 A. Kolodny, I. Kidron: Infrared Phys. **22**, 9 (1992)
- 36.10 N. Oda: Proc. SPIE **915**, 20 (1988)
- 36.11 M. A. Kinch, S. R. Borello, A. Simmons: Infrared Phys. **17**, 127 (1977)
- 36.12 D. L. Smith: J. Appl. Phys. **54**, 5441 (1983)
- 36.13 T. Ashley, C. T. Elliott: Infrared Phys. **22**, 367 (1982)
- 36.14 D. L. Smith, D. K. Arch, R. A. Wood, M. W. Scott: Appl. Phys. Lett. **45**(1), 83 (1984)
- 36.15 C. A. Musca, J. F. Siliquini, B. D. Nener, L. Faraone: IEEE Trans. Electron. Dev. **44**(2), 239 (1997)
- 36.16 R. Kumar, S. Gupta, V. Gopal, K. C. Chabra: Infrared Phys. **31**(1), 101 (1991)
- 36.17 M. A. Kinch, S. R. Borello, B. H. Breazale, A. Simmons: Infrared Phys. **16**, 137 (1977)
- 36.18 M. B. Reine, E. E. Krueger, P. O'Dette et al.: Proc. SPIE **2816**, 120 (1996)
- 36.19 I. M. Baker, F. A. Capocci, D. E. Charlton, J. T. M. Wotherspoon: Solid-State Electron. **21**, 1475 (1978)
- 36.20 C. T. Elliott: Electron. Lett. **17**, 312 (1981)
- 36.21 C. T. Elliott, N. T. Gordon, R. S. Hall, G. J. Crimes: J. Vac. Sci. Technol. A **8**, 1251 (1990)
- 36.22 C. T. Elliott: UK Patent 1488, p. 258 (1977)
- 36.23 C. T. Elliott, D. Day, D. J. Wilson: Infrared Phys. **22**, 31 (1982)

- 36.24 C. T. Elliott, C. L. Jones: *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications* (Chapman Hall, New York 1997) Chap. 16
- 36.25 A. Blackburn, M. V. Blackman et al.: *Infrared Phys.* **22**, 57 (1982)
- 36.26 D. J. Day, T. J. Shepherd: *Solid-State Electron.* **25(6)**, 707 (1982)
- 36.27 T. J. Shepherd, D. J. Day: *Solid-State Electron.* **25(6)**, 713 (1982)
- 36.28 T. Ashley, C. T. Elliott, A. M. White et al.: *Infrared Phys.* **24(1)**, 25 (1984)
- 36.29 C. T. Elliott: *Proc. SPIE* **1038**, 2 (1989)
- 36.30 A. Campbell, C. T. Elliott, A. M. White: *Infrared Phys.* **27(2)**, 125 (1987)
- 36.31 A. Sher, A. B. Chen, W. E. Spicer, C. K. Shih: *J. Vac. Sci. Technol. A* **3**, 105 (1985)
- 36.32 A. Rogalski: *Infrared Detectors and Emitters: Materials and Devices*, Electron. Mater. Vol. 8 (Kluwer Academic, Dordrecht 2001) Chap. 12
- 36.33 E. J. Smith, T. Tung, S. Sen et al.: *J. Vac. Sci. Technol. A* **5**, 3043 (1987)
- 36.34 J. Piotrowski, T. Niedziela: *Infrared Phys.* **30**, 113 (1990)
- 36.35 E. A. Patten, M. H. Kalisher, G. R. Chapman et al.: *J. Vac. Sci. Technol. B* **9**, 1746 (1991)
- 36.36 J. Ameurlaine, A. Rousseau, T. Nguyen-Duy, R. Triboulet: *Proc. SPIE* **929**, 14 (1988)
- 36.37 D. L. Kaiser, P. Becla: *Mater. Res. Soc. Symp. Proc.* **90**, 397 (1987)
- 36.38 P. Becla: *J. Vac. Sci. Technol. A* **4**, 2014 (1986)
- 36.39 R. M. Broudy, V. J. Mazurczyk: *Semicond. Semimet.*, 18 (1991)
- 36.40 M. B. Reine: *Proc. SPIE* **443**, 2 (1983)
- 36.41 M. B. Reine, K. R. Maschoff, S. B. Tobin et al.: *Semicond. Sci. Technol.* **8**, 788 (1993)
- 36.42 L. J. Kozlowski: *Proc. SPIE* **2745**, 2 (1996)
- 36.43 L. J. Kozlowski, J. Montroy, K. Vural, W. E. Klein-hans: *Proc. SPIE* **3436**, 162 (1998)
- 36.44 M. B. Reine, A. K. Sood, T. J. Tredwell et al.: *Semiconductors and Semimetals*, Vol. 18, ed. by R. K. Willardson, A. C. Beer (Academic, New York 1981) Chap. 6
- 36.45 M. B. Reine: *Infrared Detectors and Emitters: Materials and Devices*, Electron. Mater. Vol. 8 (Kluwer Academic, Dordrecht 2001) Chap. 12, p. 8
- 36.46 D. E. Lacklison, P. Capper et al.: *Semicond. Sci. Technol.* **2**, 33 (1987)
- 36.47 P. L. Polla, R. L. Aggarwal, D. A. Nelson et al.: *Appl. Phys. Lett.* **43**, 941 (1983)
- 36.48 O. K. Wu, G. S. Kamath, W. A. Radford et al.: *J. Vac. Sci. Technol. A* **8(2)**, 1034 (1990)
- 36.49 O. P. Agnihotri, C. A. Musca, L. Faraone: *Semicond. Sci. Technol.* **13**, 839–845 (1998)
- 36.50 W. W. Anderson: *Infrared Phys.* **20**, 353 (1980)
- 36.51 J. Y. Wong: *IEEE Trans. Electron. Dev.* **27**, 48 (1980)
- 36.52 W. W. Anderson, K. J. Hoffman: *J. Appl. Phys.* **53**, 9130 (1982)
- 36.53 C. T. Sah: *Phys. Rev.* **123**, 1594 (1961)
- 36.54 R. E. DeWames, J. G. Pasko, E. S. Yao, A. H. B. Van-derwyck, G. M. Williams: *J. Vac. Sci. Technol.* **A6**, 2655 (1988)
- 36.55 Y. Nemirovsky, D. Rosenfeld, R. Adar, A. Kornfeld: *J. Vac. Sci. Technol.* **A7**, 528 (1989)
- 36.56 D. Rosenfeld, G. Bahir: *IEEE Trans. Electron. Dev.* **39**, 1638–45 (1992)
- 36.57 Y. Nemirovsky, R. Fastow, M. Meyassed, A. Unikovsky: *J. Vac. Sci. Technol.* **B9(3)**, 1829 (1991)
- 36.58 I. M. Baker, C. D. Maxey: *J. Electron. Mater.* **30(6)**, 682 (2003)
- 36.59 I. M. Baker, G. J. Crimes, C. K. Ard et al.: *IEE Conf. Pub.* **321**, 78 (1990)
- 36.60 Y. Nemirovsky, A. Unikovsky: *J. Vac. Sci. Technol.* **B10**, 1602 (1992)
- 36.61 J. H. Tregilgas: *J. Vac. Sci. Technol.* **21**, 208 (1982)
- 36.62 J. P. Hirth, H. Ehrenreich: *J. Vac. Sci. Technol.* **A3**, 367 (1985)
- 36.63 A. Szilagyi, M. N. Grimbergen: *J. Cryst. Growth* **86**, 912 (1988)
- 36.64 A. J. Syllaios, L. Colombo: *Proc. IEDM Conf.* (IEEE, New York 1982) p. p137
- 36.65 B. Pellicari, G. Baret: *J. Appl. Phys.* **62**, 3986 (1987)
- 36.66 S. M. Johnson, D. R. Rhiger, J. P. Rosberg et al.: *J. Vac. Sci. Technol.* **B10**, 1499 (1992)
- 36.67 P. W. Norton, A. P. Erwin: *J. Vac. Sci. Technol.* **A7**, 503 (1989)
- 36.68 P. S. Wijewarnasuriya, M. Zandian, D. B. Young et al.: *J. Electron. Mater.* **28**, 649–53 (1999)
- 36.69 I. M. Baker, G. J. Crimes, J. E. Parsons, E. S. O'Keefe: *Proc. SPIE* **2269**, 636 (1994)
- 36.70 S. M. Johnson, J. A. Vigil, J. B. James et al.: *J. Electron. Mater.* **22**, 835 (1993)
- 36.71 S. J. C. Irvine: *Narrow-gap II–IV Compounds for Optoelectronic and Electromagnetic Applications* (Chapman and Hall, New York 1997) Chap. 3
- 36.72 J. Tunnicliffe, J. Irvine, S. Dosser, J. Mullin: *J. Cryst. Growth* **68**, 245 (1984)
- 36.73 C. D. Maxey, J. P. Camplin, I. T. Guilfooy et al.: *J. Electron. Mater.* **32(7)**, p656 (2003)
- 36.74 O. K. Wu, T. J. deLyon, R. D. Rajavel, J. E. Jensen: *Narrow-Gap II–IV Compounds for Optoelectronic and Electromagnetic Applications*, Part 1 (Chapman and Hall, New York 1997) Chap. 4
- 36.75 M. V. Blackman et al.: *Elect. Lett.* **23**, 978 (1987)
- 36.76 S. Margalit, Y. Nemirovsky, I. Rotstein: *J. Appl. Phys.* **50**, 6386 (1979)
- 36.77 A. Kolodny, I. Kidron: *IEEE Trans. Electron. Dev.* **ED-27**, 37 (1980)
- 36.78 L. O. Bubulac, W. E. Tennant, R. A. Riedel et al.: *J. Vac. Sci. Technol.* **21**, 251 (1982)
- 36.79 L. O. Bubulac, W. E. Tennant et al.: *Appl. Phys. Lett.* **51**, 355 (1987)
- 36.80 J. Syz, J. D. Beck, T. W. Orient, H. F. Schaake: *J. Vac. Sci. Technol.* **A7**, 396 (1989)
- 36.81 M. A. Kinch: *Proc. SPIE* **4369**, 566 (2001)
- 36.82 J. White et al.: *J. Electron. Mater.* **30(6)**, 762 (2001)

- 36.83 L. O. Bubulac, C. R. Viswanathan et al.: J. Cryst. Growth **123**, 555 (1992)
- 36.84 I. M. Baker, R. A. Ballingall: Proc. SPIE. **510**, 210 (1985)
- 36.85 P. Tribulet, J.-P. Chatard, P. Costa, S. Paltrier: J. Electron. Mater. **30**(6), 574 (2001)
- 36.86 T. Tung, M. H. Kalisher, M. H. Stevens et al.: Mater. Res. Soc. Symp. Proc. **90**, 321 (1987)
- 36.87 C. C. Wang: J. Vac. Sci. Technol. **B9**, 740 (1991)
- 36.88 G. N. Pulz, P. W. Norton, E. E. Krueger, M. B. Reine: J. Vac. Sci. Technol. **B9**, 1724 (1991)
- 36.89 P. W. Norton, P. LoVecchio, G. N. Pultz et al.: Proc. SPIE **2228**, 73 (1994)
- 36.90 T. Tung: J. Cryst. Growth **86**, 161 (1988)
- 36.91 J. Arias, M. Zandian, J. G. Pasko et al.: J. Appl. Phys. **69**, 2143 (1991)
- 36.92 J. M. Arias, J. G. Pasko, M. Zandian et al.: Appl. Phys. Letts. **62**, 976 (1993)
- 36.93 J. Bajaj: Proc. SPIE **3948**, 42 (2000)
- 36.94 K. W. Hodapp, J. K. Hora, D. N. B. Hall et al.: New Astronomy **1**, 177 (1996)
- 36.95 C. D. Maxey, C. J. Jones, N. Metcalf et al.: Proc. SPIE **3122**, 453 (1996)
- 36.96 J. B. Varesi, R. E. Bornfreund, A. C. Childs et al.: J. Electron. Mater. **30**(6), 56698 (2001)
- 36.97 D. J. Hall, L. Buckle, N. T. Gordon et al.: Proc. SPIE **5406**, 317 (2004)
- 36.98 G. Bostrup, K. L. Hess, J. Ellsworth, D. Cooper, R. Haines: J. Electron. Mater. **30**(6), 560 (2001)
- 36.99 K. Vural, L. J. Kozlowski, D. E. Cooper et al.: Proc. SPIE **3698**, 24 (1999)
- 36.100 N. T. Gordon, I. M. Baker: *Infrared Detectors and Emitters: Materials and Devices*, Electron. Mater. Vol. 8 (Kluwer Academic, Dordrecht 2001) Chap. 2, p. 23
- 36.101 J. D. Beck, C.-F. Wan, M. A. Kinch, J. E. Robinson: Proc. SPIE **4454**, 188 (2001)
- 36.102 M. A. Kinch, J. D. Beck, C.-F. Wan et al.: J. Electron. Mater. **33**(6), 630 (2003)
- 36.103 T. J. de Lyon, J. E. Jenson, M. D. Gordwitz et al.: J. Electron. Mater. **28**, 705 (1999)
- 36.104 I. M. Baker, S. S. Duncan, J. W. Copley: Proc. SPIE **5406**, 133 (2004)
- 36.105 M. B. Reine, A. Hairston, P. O'Dette et al.: Proc. SPIE **3379**, 200 (1998)
- 36.106 W. E. Tennant, M. Thomas, L. J. Kozlowski et al.: J. Electron. Mater. **30**(6), 590 (2001)
- 36.107 W. Cabanski, R. Brieter, R. Koch et al.: Proc. SPIE **4369**, 547 (2001)
- 36.108 J. M. Arias, M. Zandian, G. M. Williams: J. Appl. Phys. **70**(8), 4620 (1991)
- 36.109 R. D. Rajavel, D. M. Jamba, O. K. Wu et al.: J. Electron. Mater. **26**, 476 (1997)
- 36.110 R. D. Rajavel, D. M. Jamba, O. K. Wu et al.: J. Electron. Mater. **27**, 747 (1998)
- 36.111 C. T. Elliott, N. T. Gordon, A. M. White: Appl. Phys. Lett. **74**, 2881 (1999)
- 36.112 C. T. Elliott: *Infrared Detectors and Emitters: Materials and Devices*, Electron. Mater. Vol. 8 (Kluwer Academic, Dordrecht 2001) Chap. 11
- 36.113 N. T. Gordon, C. L. Jones, D. J. Lees et al.: Proc. SPIE **5406**, 145 (2004)
- 36.114 C. T. Elliott, T. Ashley: Electron. Lett. **21**, 451 (1985)
- 36.115 D. Marr: *Vision* (W. H. Freeman, San Francisco 1982)
- 36.116 M. Masie, P. McCarley, J. P. Curzan: Proc. SPIE **1961**, 17 (1993)
- 36.117 P. McCarley: Proc. SPIE **3698**, 716 (1999)
- 36.118 C. R. Baxter, M. A. Massie, P. L. McCarley, M. E. Couture: Proc. SPIE **4369**, 129 (2001)
- 36.119 N. T. Gordon: *Narrow-Gap II-IV Compounds for Optoelectronic and Electromagnetic Applications* (Chapman and Hall, New York 1997) Chap. 17
- 36.120 P. Bouchut, G. Destefanis, J. P. Chamonal et al.: J. Vac. Sci. Technol. B **9**, 1794 (1991)
- 36.121 T. Ashley, C. T. Elliott, N. T. Gordon et al.: Infrared Phys. Technol. **36**, 1037 (1995)
- 36.122 R. Zucca, J. Bajaj, E. R. Blazewski: J. Vac. Sci. Technol. A **6**, 2725 (1988)
- 36.123 J. R. Meyer, I. Vurgaftman: *Infrared Detectors and Emitters: Materials and Devices*, Electron. Mater. Vol. 8 (Kluwer Academic, Dordrecht 2001) Chap. 14

Optoelectronic

37. Optoelectronic Devices and Materials

Unlike the majority of electronic devices, which are silicon based, optoelectronic devices are predominantly made using III–V semiconductor compounds such as GaAs, InP, GaN and GaSb and their alloys due to their direct band gap. Understanding the properties of these materials has been of vital importance in the development of optoelectronic devices. Since the first demonstration of a semiconductor laser in the early 1960s, optoelectronic devices have been produced in their millions, pervading our everyday lives in communications, computing, entertainment, lighting and medicine. It is perhaps their use in optical-fibre communications that has had the greatest impact on humankind, enabling high-quality and inexpensive voice and data transmission across the globe. Optical communications spawned a number of developments in optoelectronics, leading to devices such as vertical-cavity surface-emitting lasers, semiconductor optical amplifiers, optical modulators and avalanche photodiodes. In this chapter we discuss the underlying theory of operation of the most important optoelectronic devices. The influence of carrier–photon interactions is discussed in the context of producing efficient emitters and detectors. Finally we discuss how the semiconductor band structure can be manipulated to enhance device properties using quantum confinement and strain effects, and how the addition of dilute amounts of elements such as nitrogen is having a profound effect on the next generation of optoelectronic devices.

37.1	Introduction to Optoelectronic Devices ..	888
37.1.1	Historical Perspective.....	888
37.2	Light-Emitting Diodes and Semiconductor Lasers	890
37.2.1	Carrier–Photon Interactions in Semiconductors.....	890
37.2.2	Direct- and Indirect-Gap Semiconductors	890
37.2.3	Emission and Absorption Rates and the Einstein Relations	891
37.2.4	Population Inversion	892
37.2.5	Gain in Semiconductors	892
37.2.6	Density of States	893
37.2.7	Optical Feedback in a Fabry–Perot Laser Cavity	896
37.2.8	Wave-Guiding	897
37.2.9	Carrier Confinement	898
37.2.10	Current Confinement	898
37.2.11	Laser Threshold and Efficiency	899
37.2.12	Carrier Recombination Processes .	900
37.2.13	Temperature Sensitivity and T_0 ...	903
37.3	Single-Mode Lasers.....	904
37.3.1	DFB lasers	904
37.3.2	VCSELs	905
37.4	Optical Amplifiers	906
37.4.1	An Introduction to Optical Amplification	906
37.4.2	Semiconductor Optical Amplifiers (SOAs)	907
37.5	Modulators	907
37.5.1	Modulator Theory.....	907
37.5.2	Polarisation-Insensitive Modulators.....	909
37.5.3	High-Speed High-Power QCSE Modulators.....	910
37.5.4	The Electro-Optic Effect.....	911
37.6	Photodetectors.....	911
37.6.1	Photodetector Requirements	912
37.6.2	Photodetection Theory.....	912
37.6.3	Detectors with Internal Gain.....	913
37.6.4	Avalanche Photodetectors	913
37.7	Conclusions	914
	References	915

37.1 Introduction to Optoelectronic Devices

In this chapter we introduce the underlying theory and operating principles of semiconductor optoelectronic devices. There exist today a plethora of optoelectronic devices, which are used in a multitude of applications. These devices include sources such as light-emitting diodes (LEDs) and laser diodes, photodetectors, optical amplifiers and optical modulators. With such devices, one can generate, modulate, detect and switch photons in an analogous way to electrons in an electrical circuit. We begin this chapter by considering the underlying physical interactions between electrons and photons that occur in semiconductors and how they may be harnessed to produce a wide variety of devices. At the time of writing, optoelectronic devices have found their way into many different aspects of modern life whether it be the ubiquitous indicator LEDs on hi-fi systems, televisions, computers, solid-state lighting and countless other items or in the bar-code scanning systems at the supermarket, the compact disc (CD) player, CD-ROM/CD±R/CD±RW or digital versatile disk (DVD), DVD-ROM/DVD±R/DVD±RW/DVD-RAM at home, the laser printer in the office, or when using a telephone or watching cable television. Over the past decade or so there has been an information explosion whereby information from all over the world can be quickly accessed by anyone equipped with a computer and access to the internet. In all of these applications it is a semiconductor-based optoelectronic device that forms an essential part of the system.

One of the major advantages of semiconductor devices is their small size. For example, a typical edge-emitting laser measures approximately 500 μm long by 250 μm wide with a thickness of 100 μm . Several thousand such devices can be made from a single wafer. Thus, even when packaged, these form very compact sources of coherent radiation. Other types of laser, such as the gas laser, simply cannot compare with the semiconductor laser in terms of size, modulation rates, flexibility of application and power consumption. Furthermore, semiconductor devices can be tailored to meet the exacting requirements of an application by simply altering the composition of the various layers forming the structure.

Of the uses of semiconductor devices outlined above, their use in telecommunications stands out as having the largest impact on modern life. Digital-based data transmission allows information to be transmitted over large distances with a much lower degradation in signal quality compared with older, analogue-based

systems. Optical telecommunications are ideal for use in digital systems, enabling data transmission rates in excess of 10 Gbit/s using short optical pulses of < 100 ps in duration. Even at this exceptionally fast rate of data transmission, bit error rates of better than 1 in every 10^9 bits can be achieved. Such capabilities are a direct consequence of significant research and development work that has gone into producing semiconductor devices for the emission and detection of light. In the next section, we give a brief history of this development before proceeding to discuss the key elements of physics and technology related to the device operation.

37.1.1 Historical Perspective

The development of optoelectronic devices began in the early 1960s with the development of the light-emitting diode and soon thereafter, the semiconductor laser. *Holonyak* had been experimenting with the alloy GaAsP to produce visible light via spontaneous emission and was successful in producing visible (red) light-emitting diodes [37.1]. Stimulated emission was first predicted by *Albert Einstein* in his famous 1917 paper, *Zur Quantentheorie der Strahlung* (On the quantum theory of radiation), [37.2]. In 1961 the possibility of obtaining stimulated emission in semiconductors was discussed by *Bernard* and *Duraffourg* who, for the first time, derived the condition for lasing action in semiconductor materials [37.3]. This became known as the Bernard–Duraffourg condition, as discussed later in this chapter. In the following year, the first reports of lasing action in semiconductor materials were published by four independent groups, *Nathan* and co-workers [37.4], *Hall* and co-workers [37.5], *Quist* and co-workers [37.6] and *Holonyak* and co-workers [37.1]. These lasers were based upon GaAs which, with its direct band gap, made it suitable for use as an optical source. The first devices consisted of simple p–n homo-junctions which have a single interface between the n- and p-doped regions. With a wavelength of near to 900 nm, these lasers emitted in the near-infrared region of the electromagnetic spectrum. Following on from his success with LEDs, *Holonyak* produced visible (red) semiconductor lasers. The first semiconductor lasers had very high threshold current densities, J_{th} (defined later in this chapter), and could only be operated under pulsed conditions. This made such devices impractical but started a period of intensive research into produc-

ing the first continuous-wave (CW) room-temperature semiconductor-based laser. After much investment into developing the growth technology of liquid-phase epitaxy (LPE), it became possible to produce a high-quality double heterostructure consisting of GaAs sandwiched between higher-band-gap $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layers. The double heterostructure brought two key advantages over the homojunction; firstly, the lower-band-gap GaAs region formed a reservoir for the carriers (carrier confinement) where they could recombine across the band gap, and secondly, the higher refractive index of GaAs with respect to $\text{Al}_x\text{Ga}_{1-x}\text{As}$ provided better confinement of the optical field. These two improvements resulted in a significant reduction in J_{th} of two orders of magnitude and enabled CW operation at room temperature to be achieved for the first time [37.7, 8]. Almost 30 years later, in 2000, Alferov and Kroemer were co-recipients of the Nobel prize for physics for their pioneering work on the development of semiconductor heterostructures. In the late 1970s further improvements in semiconductor growth technology led to the development of molecular-beam epitaxy (MBE) and vapour-phase epitaxy (VPE) which enabled increasingly thin layers to be grown reproducibly. Layer thicknesses of the order of less than 100 \AA became achievable and introduced the regime in which quantum-confinement effects could be harnessed. These quantum well (QW) structures [37.9] brought about further improvements in laser performance, including a further increase in carrier confinement, narrower line width and extended wavelength tunability for a given material composition. The splitting of the degeneracy of the valence band also resulted in a reduced density of states at the top of the valence band. These improvements led to a substantially reduced J_{th} . In 1986, Adams, and also Yablonoitch and Kane, independently predicted that the introduction of biaxial strain into the active region of quantum wells would result in a further improvement in the laser characteristics [37.10, 11]. The proposed benefits of strain included a further reduction in the density of states at the top of the valence band (for compressive strain) and the ability to tailor the symmetry of the carrier distribution to that of the laser beam. The introduction of tensile strain made it possible to produce a semiconductor laser with transverse magnetic (TM) polarisation, something that was not previously possible. Strained-layer quantum wells are now included in the majority of commercially available semiconductor laser products. Recent material developments include quantum dot lasers, in which carriers are restricted in all three dimensions. There is a major

effort to develop quantum dot (QD) lasers with better performance than existing quantum well lasers. More than 20 years ago the first theoretical prediction [37.12] showed that using three-dimensionally (3D)-confined structures with an atomic-like discrete density of states in the active region of semiconductor laser should allow the development of devices with low threshold current density and very high thermal stability. Since this time, quantum dot lasers have been demonstrated with record low room-temperature threshold current densities, $< 20 \text{ A/cm}^2$ [37.13]. Work is ongoing to produce temperature-insensitive quantum dot lasers, particularly at $1.3 \text{ }\mu\text{m}$.

The late 1970s and early 1980s also gave rise to the development of single-mode lasers such as the distributed feedback (DFB) laser which have made long-haul optical communications possible [37.14]. The development of such advanced semiconductor laser devices spawned research into other semiconductor optoelectronic devices such as monolithic tunable lasers (discussed in detail by [37.15]), the semiconductor optical amplifier (SOA), optical modulators and advanced photodetectors. These devices are discussed in the latter part of this chapter.

In parallel with the developments mentioned above, other technologies were being investigated such as the vertical-cavity surface-emitting laser (VCSEL) originally proposed by Iga in 1977 and demonstrated by his group in 1979 (for a comprehensive review see [37.16]). VCSELs promise low-cost single-mode lasers which can easily be made into arrays. There remains intense activity to develop VCSELs for operation at the important telecommunications wavelength of $1.55 \text{ }\mu\text{m}$. VCSELs are discussed in later in this chapter. Other devices of note include the quantum cascade laser (QCL) developed by Capasso and co-workers in 1994 [37.17]. The development of QCLs was driven by the need to produce lasers emitting in the mid-infrared (mid-IR) ($\approx 3\text{--}12 \text{ }\mu\text{m}$) for gas sensing and environmental monitoring applications. For conventional interband lasers as discussed above, this requires narrow-band-gap materials which are much less developed than materials such as GaAs and InP. QCLs get around this problem by utilising *intra*-band transitions involving only one type of carrier, e.g. electrons (thus they are unipolar devices), and can be grown using conventional GaAs- or InP-based materials. In such devices, the carrier cascades through several quantum wells, giving rise to many photons per injected electron, quite unlike conventional interband devices. Clearly, QCLs offer great possibilities for the future development of mid-IR lasers

and perhaps, with suitable materials such as GaN, they may allow emission in the near-IR (for a review see [37.18]).

Unfortunately, it is impossible to go into detail about the developments of the full range of optoelectronic de-

vices in a single book chapter. However, we hope that this chapter provides a basic grounding on the relevant physics of the most important optoelectronic devices allowing the interested reader to investigate further, starting with the bibliography.

37.2 Light-Emitting Diodes and Semiconductor Lasers

37.2.1 Carrier–Photon Interactions in Semiconductors

In its simplest form, a semiconductor can be considered as a system containing two energy bands populated with electrons. In such a system there exist three possible processes by which the electrons can move between the two bands: the spontaneous emission, absorption and stimulated emission processes, as shown schematically in Fig. 37.1. Spontaneous emission occurs when an electron in the conduction band with energy E_2 recombines with a hole in the valence band at a lower energy E_1 . The difference in energy, $(E_2 - E_1)$ is emitted in the form of a photon. This process, as its name implies, is random and photons may be emitted in any direction with arbitrary polarisation. The probability of spontaneous emission is proportional to the density of electrons in the conduction band (n) and to the density of holes in the valence band (p). In the absorption process, an electron in the valence band is promoted to the conduction band by the absorption of a photon with an energy equal to the optical band gap. The probability of absorption is

proportional to the photon density, the density of electrons in the valence band and the density of empty states in the conduction band. The stimulated emission process occurs when a photon of energy $E_2 - E_1$ interacts with an electron in the conduction band, causing it to recombine with a hole in the valence band thereby generating a photon. The photon released by this process has both identical energy, phase and momentum to that of the incident photon. The probability of stimulated emission is therefore proportional to n , p and the photon density. In contrast to spontaneous emission, stimulated emission produces photons that are essentially identical and therefore very pure monochromatic light is generated.

At room temperature in an undoped semiconductor, there are far fewer electrons in the conduction band than in the valence band. As a consequence, absorption is far more likely than emission. To achieve optical gain, the probability of stimulated emission must exceed that of absorption. By applying either an electrical current (electrical pumping) or by injecting light with photon energies greater than the transition energy (optical pumping), it is possible to inject electrons into the conduction band and holes into the valence band. Electrical pumping is used for practical devices.

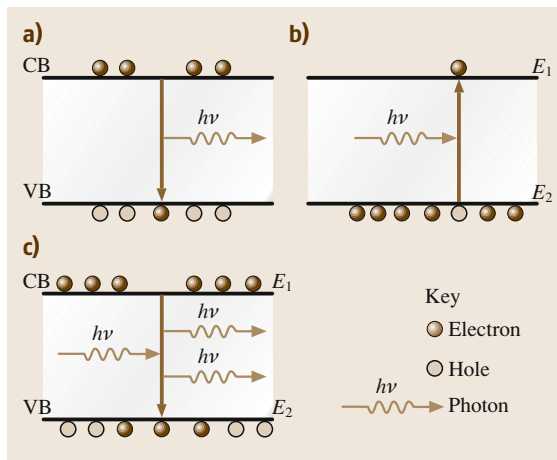


Fig. 37.1a–c The three electron–photon interactions in semiconductors: (a) spontaneous emission, (b) stimulated absorption, and (c) stimulated emission

37.2.2 Direct- and Indirect-Gap Semiconductors

Semiconductors may exist in two basic forms where the band gap is either direct or indirect. Whether or not the band gap is direct or indirect has a profound influence on their suitability for use in optoelectronic devices. Figure 37.2 illustrates the difference between a direct- (e.g. GaAs) and an indirect-band-gap (e.g. Ge) semiconductor. It can be seen that for the direct-gap material, the conduction-band (CB) energy minimum occurs at the same k -value as the valence-band (VB) maximum (direct). In contrast, for the indirect-gap material the CB minimum lies at a different k -value to the VB maximum. For any electron transition, the total energy and momentum must be conserved. The photon wavevector has a magnitude given by $2\pi/\lambda$, where λ

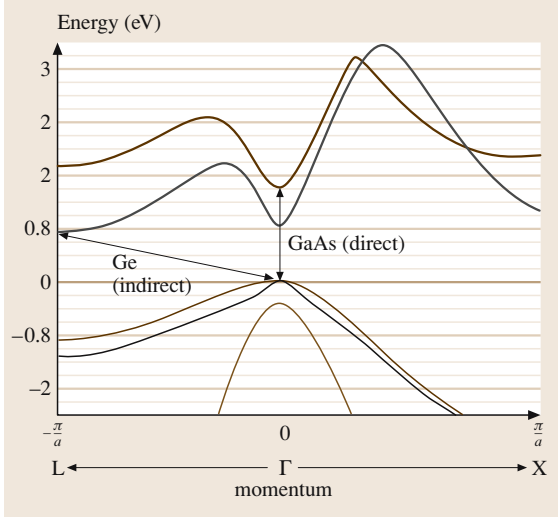


Fig. 37.2 Band structure of a direct-gap semiconductor (GaAs) and an indirect semiconductor (Ge)

is typically $\approx 1 \mu\text{m}$, whilst the magnitude of the electron wavevector ranges between $-\pi/a$ and π/a within the first Brillouin zone, where a is the crystal lattice spacing, typically $\approx 1 \text{ \AA}$. Clearly therefore, the photon wavevector is much smaller than the possible electron wavevectors and hence, if a photon interacts with an electron, the transition must occur with virtually no change in wavevector, hence only vertical transitions are allowed. It is clear that such transitions can occur in the direct-band-gap semiconductor. However, in the case of indirect semiconductors, such transitions are only possible via the interaction of a phonon. Since this becomes a three-particle interaction, the transition probability is significantly reduced. Stimulated emission is inhibited in indirect-band-gap semiconductors and it is for this reason that direct-band-gap semiconductors (such as GaAs, InP and related alloys) are primarily used to produce photoemissive devices such as LEDs and lasers. Indirect semiconductors, such as elemental germanium or silicon are therefore impractical for the production of such devices. However, these are very effective semiconductor materials for detection, as will be discussed later in this chapter.

37.2.3 Emission and Absorption Rates and the Einstein Relations

The populations of electrons and holes in the conduction and valence bands respectively are governed by both the densities of states and the Fermi–Dirac occupation

probabilities. In a semiconductor, electron–electron and hole–hole scattering rates are typically $\approx 100 \text{ fs}$ whilst electron–hole recombination times are $\approx 1 \text{ ns}$. We can therefore assume that the electrons and holes are in thermal equilibrium with themselves and can be described using a Fermi–Dirac distribution. It is useful to define two quasi-Fermi levels. These are referred to as F_c and F_v , corresponding to the energy at which the occupation probability equals $\frac{1}{2}$ for electrons in the CB and VB, respectively. In thermal equilibrium, $F_c = F_v$. The corresponding energy-dependent Fermi functions are

$$f_c(E) = \frac{1}{1 + \exp\left(\frac{E - F_c}{k_B T}\right)} \quad \text{and} \quad f_v(E) = \frac{1}{1 + \exp\left(\frac{E - F_v}{k_B T}\right)}, \quad (37.1)$$

where k_B is the Boltzmann constant and T is the absolute temperature.

By considering each of the processes illustrated in Fig. 37.1 one can obtain mathematical expressions relating each of the processes to the another. If we label states in the VB as “1” and states in the CB as “2”, the absorption rate, r_{12} , for photons of energy, $h\nu$ is given by

$$r_{12} = B_{12} \rho_v f_v (1 - f_c) \rho_c P(h\nu) \quad (37.2)$$

where ρ_c and ρ_v are the CB and VB densities of states, respectively. $P(h\nu)$ is the photon density at energy $h\nu$ and B_{12} is the Einstein coefficient for the absorption process. In a similar way, the stimulated emission rate, r_{21} , may be written as

$$r_{21} = B_{21} \rho_c f_c (1 - f_v) \rho_v P(h\nu) \quad (37.3)$$

where B_{21} is the Einstein coefficient for the stimulated emission process. For the spontaneous emission rate r_{21}^{spon}

$$r_{21}^{\text{spon}} = A_{21} \rho_c f_c (1 - f_v) \rho_v. \quad (37.4)$$

A_{21} is the Einstein coefficient for the spontaneous emission process. Note that, since spontaneous emission does not require a photon to initiate the process, r_{21}^{spon} does not depend on $P(h\nu)$. Under steady-state conditions, the total upward transition rate must equal the total downward transition rate, thus

$$r_{12} = r_{21} + r_{21}^{\text{spon}}. \quad (37.5)$$

Thus, by combining equations (37.2)–(37.4), we obtain

$$\begin{aligned} B_{12} \rho_v f_v (1 - f_c) \rho_c P(h\nu) \\ = B_{21} \rho_c f_c (1 - f_v) \rho_v P(h\nu) \\ + A_{21} \rho_c f_c (1 - f_v) \rho_v. \end{aligned} \quad (37.6)$$

In thermal equilibrium ($F_c = F_v$),

$$P(h\nu) = \frac{A_{21}}{B_{12} \exp\left(\frac{h\nu}{k_B T}\right) - B_{21}}. \quad (37.7)$$

The standard expression for black-body radiation is given by Planck's law as

$$P(h\nu) = \frac{8\pi^3 n^3 (h\nu)^2}{(hc)^3} \frac{1}{\exp\left(\frac{h\nu}{k_B T}\right) - 1}. \quad (37.8)$$

Here, n is the refractive index of the semiconductor. For simplicity, here we assume that the medium is nondispersive. From these two expressions of $P(h\nu)$ we obtain the result that

$$B_{12} = B_{21} = B \quad (37.9)$$

and

$$A_{21} = B \frac{8\pi^3}{(hc)^3} n^3 (h\nu)^2. \quad (37.10)$$

37.2.4 Population Inversion

The quantity of greatest importance to semiconductor laser operation is the *net* stimulated emission rate, which may simply be calculated as $r_{\text{stim}} = r_{21} - r_{12}$. Positive values of r_{stim} mean that an optical wave will grow in intensity as it travels through the semiconductor, whilst a negative r_{stim} implies that the optical wave would be reduced in intensity. From (37.2) and (37.3), r_{stim} can be written as

$$\begin{aligned} r_{\text{stim}} &= r_{21} - r_{12} \\ &= B_{21} \rho_c f_c (1 - f_v) \rho_v P(h\nu) \\ &\quad - B_{12} \rho_v f_v (1 - f_c) \rho_c P(h\nu) \end{aligned} \quad (37.11)$$

therefore

$$r_{\text{stim}} = r_{21} - r_{12} = B \rho_c \rho_v (f_c - f_v) P(h\nu). \quad (37.12)$$

The ratio of spontaneous emission to net stimulated emission is given by

$$\begin{aligned} \frac{r_{\text{spon}}}{r_{\text{stim}}} &= \frac{A_{21} \rho_c \rho_v f_c (1 - f_v)}{B \rho_c \rho_v f_c (f_c - f_v) P(h\nu)} \\ &= \frac{A_{21} f_c (1 - f_v)}{B P(h\nu) (f_c - f_v)}. \end{aligned} \quad (37.13)$$

Combining (37.13) with (37.10) leads to

$$\frac{r_{\text{spon}}}{r_{\text{stim}}} = \frac{8\pi^3 n^3 (h\nu)^2}{(hc)^3} \frac{f_c (1 - f_v)}{P(h\nu) (f_c - f_v)}. \quad (37.14)$$

Using the expressions for f_c and f_v (37.1), this can be simplified to

$$\frac{r_{\text{spon}}}{r_{\text{stim}}} = \frac{8\pi^3 n^3 (h\nu)^2}{(hc)^3} \frac{1}{P(h\nu) \left[1 - \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right) \right]} \quad (37.15)$$

and consequently

$$\begin{aligned} r_{\text{stim}} &= r_{\text{spon}} \frac{(hc)^3}{8\pi^3 n^3 (h\nu)^2} \\ &\quad \times \left[1 - \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right) \right]. \end{aligned} \quad (37.16)$$

From this expression we see that, when $F_c - F_v = h\nu$, $r_{\text{stim}} = 0$. Thus, when the quasi-Fermi-level splitting equals the photon energy, the absorption and stimulated emission rates cancel. At this injection level, the semiconductor is effectively transparent. This is known as the Bernard–Duraffourg condition. For photons of energy $h\nu > F_c - F_v$, there will be absorption. However, for photons of energy $h\nu < F_c - F_v$ there is gain. This defines the condition necessary to achieve population inversion in the semiconductor and lasing action can occur when $F_c - F_v > E_g$.

37.2.5 Gain in Semiconductors

If $\alpha(h\nu)$ is the rate of gain/loss per unit length and $P(h\nu)$ is the photon density, then the rate at which the photon density increases/decreases per unit length is simply $\alpha(h\nu)P(h\nu)$. The corresponding rate of increase of photon density per unit time is then

$$r_{\text{stim}} = P(h\nu) \alpha(h\nu) \frac{c}{n} \quad (37.17)$$

where c/n is the speed of light within a semiconductor of refractive index, n . By equating our two expressions for r_{stim} [(37.16) and (37.17)] we find that

$$\begin{aligned} \alpha(h\nu) &= r_{\text{spon}} \frac{h^3 c^2}{8\pi^3 n^2 (h\nu)^2} \\ &\quad \times \left[1 - \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right) \right]. \end{aligned} \quad (37.18)$$

Under conditions of low pumping or at high photon energies ($h\nu \gg F_c - F_v$), (37.18) can be approximated to

$$\begin{aligned} \alpha(h\nu) &= -r_{\text{spon}} \frac{h^3 c^2}{8\pi^3 n^2 (h\nu)^2} \\ &\quad \times \exp\left(\frac{h\nu - (F_c - F_v)}{k_B T}\right) \end{aligned} \quad (37.19)$$

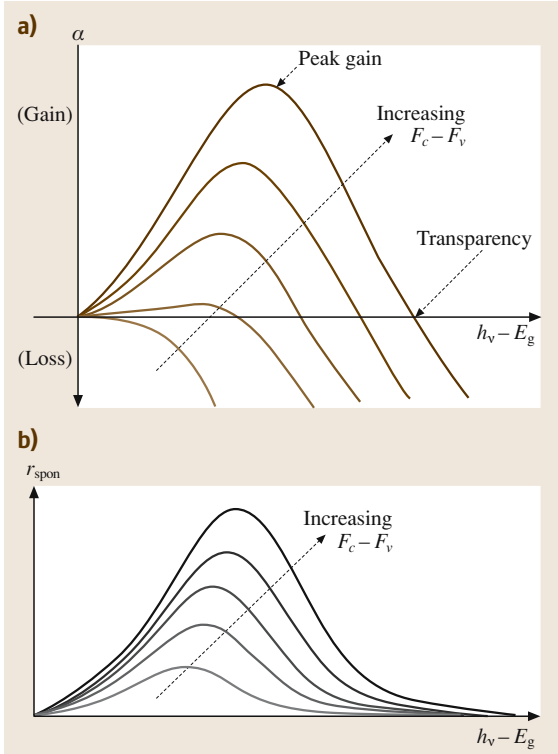


Fig. 37.3 (a) gain/loss and (b) spontaneous emission spectra as a function of increasing quasi-Fermi-level splitting

from which we have

$$r_{\text{spon}} = -\alpha(h\nu) \frac{8\pi^3 n^2 (h\nu)^2}{h^3 c^2} \times \exp\left(\frac{(F_c - F_v) - h\nu}{k_B T}\right) \quad (37.20)$$

Typical plots of the absorption and spontaneous emission rates can be seen in Fig. 37.3. From Fig. 37.3a it can be seen that, with increasing pumping level (increasing quasi-Fermi-level splitting), the semiconductor moves from purely absorption and becomes a gain medium. With increasing pumping, the spread of photon energies over which gain occurs also increases, with the peak gain moving towards higher photon energy. In Fig. 37.3b, the spontaneous emission rate is plotted as a function of photon energy. From this it can be seen that the emission spectrum is broad and corresponds to the output spectrum of a light-emitting diode (LED). With increasing pumping level, the integrated spontaneous emission increases whilst the peak shifts to higher photon energies. In practical devices, current-induced Joule heating

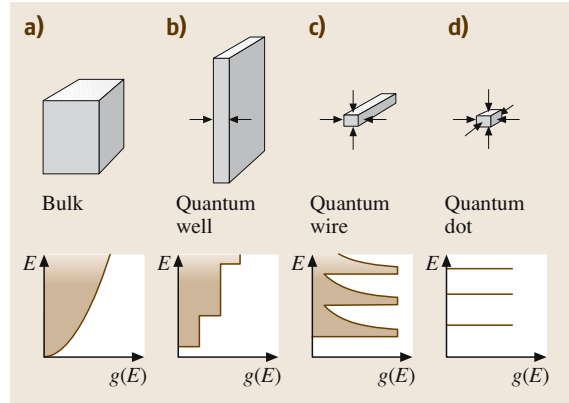


Fig. 37.4a-d Density of states in systems of differing dimensionality: (a) bulk, (b) quantum well, (c) quantum wire, and (d) quantum dot

may lead to a slight decrease in the band gap, which reduces the overall blue shift.

37.2.6 Density of States

As a semiconductor laser is put under forward bias, the quasi-Fermi levels in the conduction and valence bands move towards higher energies. The separation of the quasi-Fermi levels is dependent on the density of states $\rho(E)$ in the conduction and valence bands. At laser transparency, where the quasi-Fermi-level separation equals the photon energy, the transparency carrier density, n_{tr} and hence the threshold carrier density, n_{th} depend upon the number of states in a given energy range. By reducing the number of available states, n_{th} and consequently I_{th} (J_{th}), can be lowered. The density of states can be modified by changing the band structure. This can be achieved by reducing the dimensionality, thereby limiting the motion of the carriers, and also by introducing strain into the active region of the laser, as shall be described shortly.

Dimensionality

In conventional bulk semiconductor materials, the carriers have unrestricted motion in all directions and can be described by a 3D density of states,

$$\rho_{3\text{-D}} \propto m_{\text{eff}}^{3/2} \sqrt{E} \quad (37.21)$$

where m_{eff} is the effective mass. This is illustrated in Fig. 37.4a, where the shaded region indicates the carrier population as a function of energy. In such (bulk) devices there are relatively few states near to the band edges and hence the quasi-Fermi levels move towards

higher energies (where there is a larger concentration of available states) in order to achieve sufficient gain. This results in a broad output spectrum corresponding to radiative transitions over a wide energy range. By restricting the motion of the carriers along one dimension, as in quantum well structures, the density of states becomes independent of energy, with many states at the band edges. In a quantum well, the density of states is dependent only on the effective mass, m_{eff} and the quantum well width, L_z as:

$$\rho_{2\text{-D}} \propto \frac{m_{\text{eff}}}{L_z}. \quad (37.22)$$

In bulk structures, at the Brillouin zone centre of the valence band, the light- and heavy-hole bands are degenerate. Consequently, as the laser is pumped, states that do not contribute to the gain of the lasing mode also become filled. In a quantum well, the energy states form discrete levels. For an infinite square well, the confinement energy of the n^{th} energy level E_n is dependent on the quantum well width L_z and the growth-direction effective mass m_{eff}^z , and at the zone centre can be simply expressed analytically as

$$E_n = \frac{\hbar^2}{2m_{\text{eff}}^z} \left(\frac{n\pi}{L_z} \right)^2. \quad (37.23)$$

Thus, in a quantum well, the degenerate heavy- and light-hole bands become split (due to the difference in m_{eff}^z), resulting in a further reduction in the density of states at the valence-band Brillouin zone centre. Therefore in a quantum well laser, a higher proportion of the filled states can contribute to the gain of the lasing mode.

Further improvements in carrier confinement can be achieved by moving towards quantum wire or even quantum dot structures. In quantum wire lasers, the carriers are further confined, allowing freedom of movement along one direction only (Fig. 37.4c). In principle, such structures can be realised, e.g. by growth along lithographically defined V-grooves and lasers have been successfully produced in this manner [37.19] although the complexity of the processing has meant that they have yet to become a commercial product. In a quantum dot laser the carriers are confined still further so that they are restricted in all three dimensions (Fig. 37.4d). Thus, an ideal quantum dot is analogous to an atom. By completely restricting the carrier motion, one may theoretically expect to achieve temperature-insensitive operation [37.12]. Initial attempts to produce quantum dot lasers concentrated on lithographic processes but it was not until the successful demonstration of self-assembled growth (the

Stranski–Krastanov technique) that quantum dot lasers became a reality. Record low room-temperature threshold current densities ($< 20 \text{ A/cm}^2$) have been achieved using quantum dot lasers (e.g. [37.13]). However, due to the low gain, long lasers are required which increases the threshold current.

The influence of strain

A detailed discussion of strain in semiconductors is given in Chapt. 20 of this book. However, due to the enormous importance of strain in optoelectronic devices, in this section, we discuss the influence of strain on device characteristics. For many years, manufacturers of semiconductor lasers struggled to produce dislocation- and strain-free material. In 1986, it was suggested independently by *Adams* and by *Yablonovitch* and *Kane* that the incorporation of strain into the active layer of semiconductor lasers would lead to improved performance [37.10, 11]. Many of these benefits have since been demonstrated [37.20–22]. The majority of the lasers currently manufactured in the world today are based upon strained layers.

In a strained-layer structure, the lattice constant of the epilayer material a_{epi} is different from the lattice constant of the substrate a_{sub} on which it is grown. When a thin epilayer of this material is grown onto the much thicker substrate, the lattice constant of the grown epilayer becomes equal to that of the substrate, a_{sub} , and the epilayer is in a state of biaxial stress. The resulting net strain in the (x – y) plane of the layer ε is given by

$$\varepsilon_{\parallel} = \varepsilon_{xx} = \varepsilon_{yy} = (a_{\text{sub}} - a_{\text{epi}})/a_{\text{epi}}, \quad (37.24)$$

where it is assumed that the substrate is much thicker than the epilayer. Hence $a_{\text{epi}} > a_{\text{sub}}$ results in compressive strain while $a_{\text{epi}} < a_{\text{sub}}$ produces tensile strain.

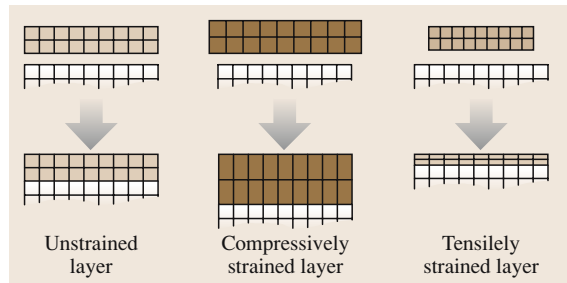


Fig. 37.5 Illustration of unstrained, compressively strained, and tensile-strained layers. For the strained layers the epilayer expands (contracts) in the plane whilst expanding (contracting) in the growth direction, corresponding to compressive (tensile) strain respectively

Figure 37.5 shows the three possible types of epilayer growth: (a) compressively strained, (b) tensile strained and (c) unstrained. The sign convention used here is that positive strain indicates that the layer is under biaxial compression whilst a negative strain corresponds to a layer being in a state of biaxial tension. For compressive strain, when the lattice constant of the epilayer is reduced in the plane, the lattice constant in the growth direction, a_{\perp} , is increased. The strain in the growth direction ε_{\perp} ($= \varepsilon_{zz}$) is governed by Poisson's ratio σ and is given by

$$\varepsilon_{\perp} = \varepsilon_{zz} = -2\sigma\varepsilon_{\parallel}/(1 - \sigma). \quad (37.25)$$

In tetrahedral semiconductors, $\sigma \approx 1/3$ therefore, upon substituting this into (37.25), it can be seen that $\varepsilon_{\perp} \approx -\varepsilon_{\parallel}$. In order to understand how strain affects the band structure, it is useful to split the strain into an axial component ε_{ax}

$$\varepsilon_{ax} = \varepsilon_{\perp} - \varepsilon_{\parallel} \approx -2\varepsilon_{\parallel} \quad (37.26)$$

and a hydrostatic component, ε_{vol} ($= \Delta V/V$, the fractional change in volume of the epilayer)

$$\varepsilon_{vol} = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \approx \varepsilon_{\parallel}. \quad (37.27)$$

The axial and hydrostatic strain components affect the band structure as described in the next section. In direct-band-gap semiconductors, whilst strain has only a small effect on the conduction band, it greatly alters the valence-band structure.

The maximum layer thickness that can be grown under a state of strain is determined by the critical thickness h_c . Above this thickness, it becomes energetically favourable to relieve the strain energy through the formation of dislocations. Thus, in order to grow high-quality strained material it is necessary to keep the thickness of each layer below h_c . As the strain in a layer is increased, the corresponding critical thickness decreases. The maximum strain–thickness product for InGaAs/GaAs materials is $\approx 200 \text{ \AA\%}$ [37.23]. It is however possible to use alternating layers of compressive and tensile strain to produce zero-net-strain structures that can prevent the formation of dislocations. Strained-layer semiconductor lasers are often fabricated with strain-compensated wells and barriers when there are many strained quantum wells, thereby providing a large strain in the wells whilst reducing the overall strain in the structure.

The introduction of strain into the active region of semiconductor diode lasers brings about several changes in the band structure. The initial predictions of Adams and Yablonoitch and Kane concentrated on

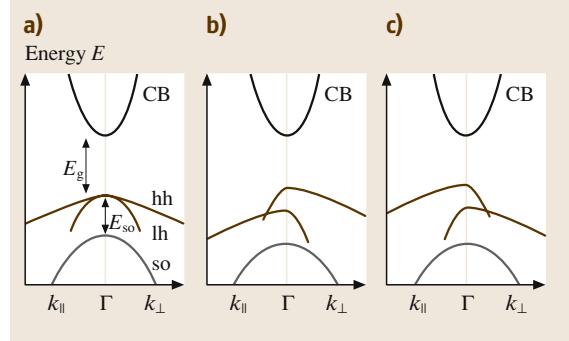


Fig. 37.6a–c Effects of strain for a bulk-like direct-band-gap semiconductor for (a) unstrained, (b) tensile strain, and (c) compressive strain

the benefits of compressive strain. It has since been shown that tensile strain can also enhance laser performance [37.24–26]. In the previous section, it was shown that the strain in an epilayer can be split into purely axial and hydrostatic components, ε_{ax} and ε_{vol} , respectively. The axial component of the strain splits the cubic symmetry of the semiconductor and hence the degeneracy of the heavy- and light-hole valence bands at the Brillouin zone centre is lifted (it should be noted that the valence band will already be nondegenerate in a quantum well due to the effects of quantum confinement). In addition, the valence-band dispersion becomes anisotropic. In the case of compressive strain the heavy-hole band is moved to higher energy with respect to the light-hole band, while the opposite is true for tensile strain. Also, for compressive strain the *in-plane* (x – y) mass becomes lighter than the mass in the growth (z) direction. Again, the opposite is true for tensile strain. The hydrostatic component of the strain causes an increase in the mean band gap for compressive strain and a decrease in the mean band gap for tensile strain. These effects on the band structure of a strained *bulk-like* material are shown schematically in Fig. 37.6.

Due to the much reduced in-plane highest hole band mass, the density of states in a compressively strained laser is lower than that for an unstrained device. As a result, n_{tr} is decreased, which not only lowers the amount of unwanted spontaneous emission [37.27] but also reduces carrier-density-dependent nonradiative processes such as Auger recombination [37.28]. Also, due to the increased heavy- and light-hole band splitting, carrier spill-over into higher sub-bands can be substantially reduced. Consequently, compressively strained lasers have been shown to have much lower threshold current densities than similar unstrained devices [37.20, 29].

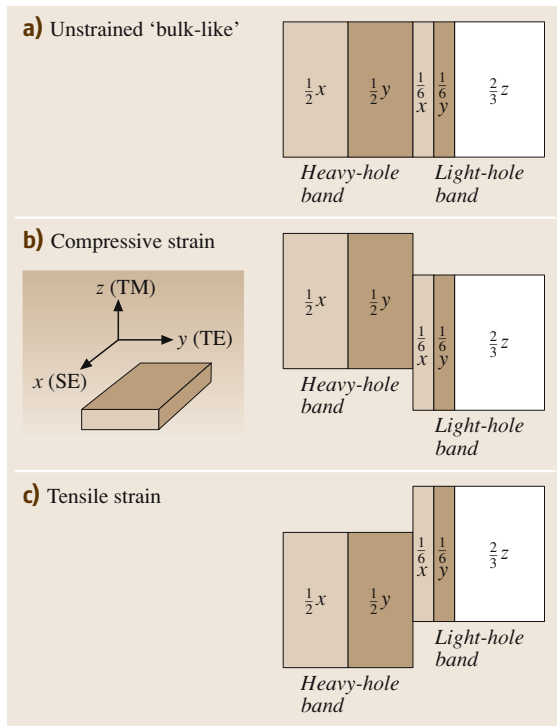


Fig. 37.7a–c Schematic illustration of the valence-band alignment for (a) unstrained bulk-like, (b) compressive strain, and (c) tensile strain

To understand the improvement in performance of tensile-strained lasers it is necessary to examine the character of the valence band. The heavy-hole band consists of $1/2$ x -like states and $1/2$ y -like states but contains no states with z -like character. In contrast, the light-hole band consists of $1/6$ x -like states, $1/6$ y -like states and $2/3$ z -like states. y -like states contribute to transverse electric (TE) gain where the electric field oscillates across the facet (y -axis) and z -like states contribute to transverse magnetic (TM) gain, where the electric field oscillates along the growth direction. x -like states contribute only to spontaneous emission (SE), as the electric field vector lies in the propagation direction of the lasing mode. This is illustrated in Fig. 37.7. In bulk, unstrained quantum well and compressively strained quantum well lasers the lasing mode is always TE-polarised as $g_{\text{th}}(\text{TE}) < g_{\text{th}}(\text{TM})$. With the inclusion of tensile strain in the quantum wells of semiconductor lasers it became possible for the first time to produce a device with TM polarisation [37.25].

From Fig. 37.6 it can be seen that the in-plane mass of a tensile-strained laser is relatively large, resulting in

an increased density of states. However, from Fig. 37.7 it can be seen that the top of the valence band in a tensile-strained device contains $2/3$ z -like states, which can contribute to TM gain. Only $1/3$ of the injected carriers will *not* produce gain in the lasing mode, compared with $2/3$ in unstrained (bulk-like) and $1/2$ in compressively strained QW devices. Hence, tensile-strained lasers can also exhibit a lower threshold current [37.25]. However, if only a moderate amount of tensile strain is incorporated into the quantum well, the threshold current will increase due to the heavy- and light-hole sub-bands becoming degenerate.

37.2.7 Optical Feedback in a Fabry–Perot Laser Cavity

The probability of stimulated emission depends not only on maintaining population inversion, but also on the photon density. In a laser, the photon density is sustained by the use of optical feedback. In a gas laser, optical feedback is realised by the use of partially metallised mirrors, which form a Fabry–Perot cavity around the gain medium. When the light reaches the end of the cavity, a certain fraction (R) is reflected back into the gain medium, where it again becomes available to stimulate the emission of further photons. In a standard edge-emitting semiconductor laser, the same principle is used but the mirrors are formed by cleaving the semiconductor along a crystallographic plane, (most commonly the $[110]$ plane) forming laser facets. This is illustrated in Fig. 37.8. By using as-cleaved facets the resulting reflectivity, R is typically $\approx 30\%$ due to the refractive-index step between the laser material and (usually) air. This can be either increased or decreased with suitable facet

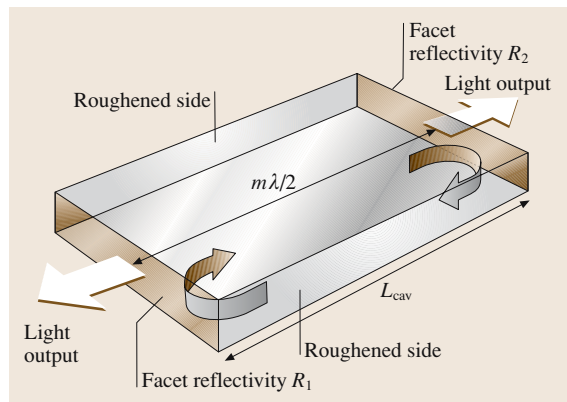


Fig. 37.8 Schematic diagram of a Fabry–Perot edge-emitting laser

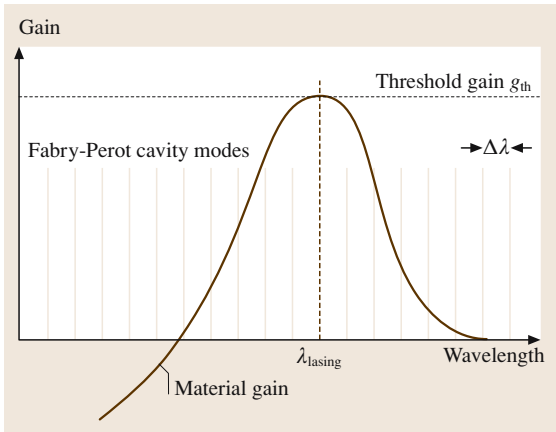


Fig. 37.9 Illustration of gain in a semiconductor laser. The cavity mode that lies closest to the material gain peak at threshold will lase

coatings. Semiconductor lasers usually have roughened sides as a result of the fabrication process. This has the desirable effect of reducing the reflectivity of the sides of the laser and hence reduces the probability of generating unwanted transverse modes.

The application of optical feedback using a Fabry–Perot cavity has the effect of producing Fabry–Perot resonances or modes in the output spectrum of the laser. From Fig. 37.8 it can be seen that the optical field will form a standing wave when the cavity length is equal to an integer number of half wavelengths, given by

$$\frac{m\lambda}{2} = nL_{\text{cav}} \quad m = 1, 2, 3, \dots \quad (37.28)$$

where m is the (integer) number of half-wavelengths, λ is the lasing wavelength (as measured), n is the refractive index of the laser cavity and L_{cav} is the cavity length (typically $250 \mu\text{m} \leq L_{\text{cav}} \leq 1500 \mu\text{m}$). It can be seen that the maximum wavelength that the cavity can support is $\lambda = 2nL_{\text{cav}}$. If (37.21) is rearranged for λ and differentiated with respect to m , the mode spacing, $\Delta\lambda$ is found to be

$$\Delta\lambda = \frac{\lambda^2}{2nL_{\text{cav}}} \quad (37.29)$$

Thus, the possible modes of a semiconductor laser consist of a comb of many hundreds of delta functions spaced $\Delta\lambda$ apart, as illustrated in Fig. 37.9. For a typical $1.5\text{-}\mu\text{m}$ laser with $L_{\text{cav}} = 500 \mu\text{m}$ and $n \approx 3.2$, $\Delta\lambda \approx 0.7 \text{ nm}$. The wavelength at which lasing occurs is therefore determined by both the material gain curve and the mode spacing. The mode which lies closest to

the peak of the material gain curve at threshold will lase (Fig. 37.9) although other modes near to the gain peak may also be present in the output spectrum of an edge-emitting Fabry–Perot laser. With increasing bias however, the relative strengths of these modes will decrease with respect to the dominant lasing mode. This produces a very narrow and spectrally pure output, in stark contrast with the spontaneous emission spectrum of an LED.

37.2.8 Wave-Guiding

As discussed earlier in this chapter, the stimulated emission process is not only dependent on the carrier density, but also on the photon density within the laser cavity. In particular, it is a requirement that the photon density is maintained as high as possible so that the probability of stimulated emission is maximised whilst still allowing sufficient radiation to escape. As the optical field propagates along the laser cavity, it can extend in both the growth (z) and transverse (y) directions. Hence it is necessary for the optical field to be confined along two directions. In a semiconductor laser, wave-guiding can be accomplished by two methods: index-guiding and gain-guiding. Index-guiding is achieved by surrounding the active region with a material of larger band gap and hence, lower refractive index. In the *growth* direction of an edge-emitting double-heterostructure semiconductor laser the optical field is always index-guided. Figure 37.10 shows an example of index-guiding in a typical double-heterostructure laser together with the associated refractive-index profile in the growth direction. Index-guiding confines most of the optical field within the active region due to total internal reflection at the cladding heterojunctions with only a small fraction

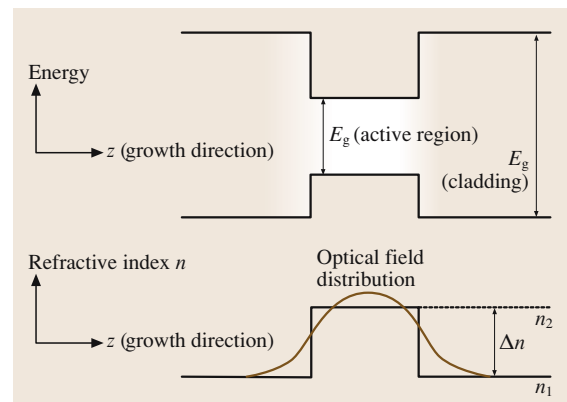


Fig. 37.10 Wave-guiding in a double heterostructure

of the field evanescently decaying into the higher-band-gap layers. The degree of confinement depends upon the refractive-index step between the active region and the surrounding layers.

In the transverse direction, the optical field can be confined by either index- or gain-guiding mechanisms. Gain-guiding arises from a weak current-induced change in refractive index across the active region caused by the nonuniform carrier distribution. Index-guided structures can be considered as either strongly index-guided or weakly index-guided, corresponding to either a large or small refractive-index step in the transverse direction. Gain-guided structures are easier to fabricate than index-guided structures but are less effective in confining the optical field when compared with index-guided structures [37.30].

37.2.9 Carrier Confinement

In addition to confining the optical mode, laser characteristics can be greatly enhanced by improving the confinement of the electrons and holes, thereby increasing the probability of radiative recombination. The first laser structures [37.1, 4–6] were based upon sin-

gle p–n junctions (homojunctions). In such structures, the injected electrons and holes recombine close to the junction. However, due to the nonuniform carrier distribution, as the electrons diffuse out of the gain region, there is a variation in gain with position. In addition, due to the small difference in refractive index at the edge of the gain region, a large amount of the optical field spreads into absorbing regions. Thus, the combination of poor carrier and field confinement gives rise to large threshold current densities $\approx 10^5 \text{ A/cm}^2$ in homojunction lasers. The development of the double-heterostructure laser [37.7, 8] brought about a large decrease in J_{th} of two orders of magnitude. This results from the greatly improved carrier confinement due to the difference in the band gap between the active and cladding layers, which forms a potential barrier to the carriers. A further improvement in carrier confinement is obtained with quantum well structures. In a quantum well laser the injected carriers rapidly diffuse into the thin ($\approx 25\text{--}150 \text{ \AA}$) quantum well layers [37.9]. Due to the large potential barriers between the well and barrier material it is difficult for the carriers to subsequently escape. In addition to improved carrier confinement, quantum well lasers offer many other benefits over bulk lasers as discussed later in this chapter. Figure 37.11 illustrates the improvement in carrier confinement obtainable on going from homojunctions to double heterostructures and quantum well laser structures.

37.2.10 Current Confinement

The threshold current of semiconductor lasers can be further decreased by reducing the area of the device that is being electrically pumped. Reducing the threshold current can also suppress current-heating effects and extend the operating lifetime of the laser. Current confinement can also help to produce a stable fundamental mode. There are several fabrication techniques and associated laser structures designed to increase current confinement [37.30]. The ridge structure Fig. 37.12a is a weakly index-guided type of laser. In this structure, following the epitaxial growth of the complete layer structure, the wafer is etched to produce a ridge that is $\approx 2\text{--}10 \mu\text{m}$ wide. The wafer is then coated with a dielectric (usually SiO_2) and metallised. The weak index-guiding is provided by the large refractive-index step between the active region material and the dielectric. The current is directly injected into the ridge and therefore travels through a relatively narrow region of the device. The two main other structures are strongly index-guided buried-

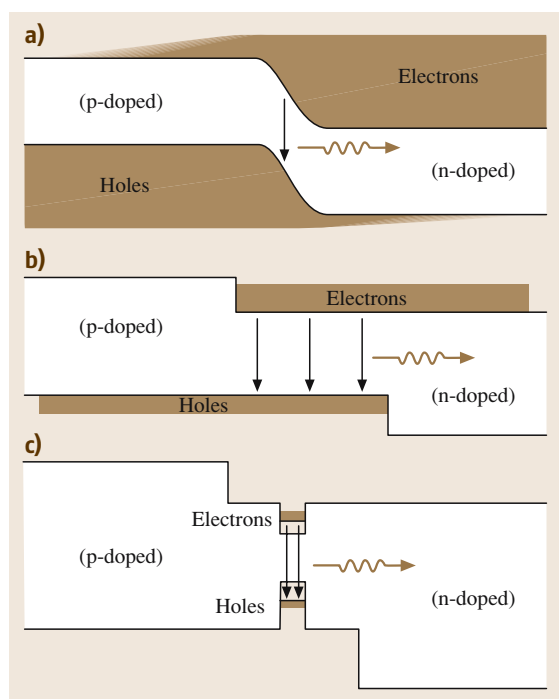


Fig. 37.11a–c Advances in carrier confinement (a) homojunction, (b) double heterostructure, and (c) quantum well

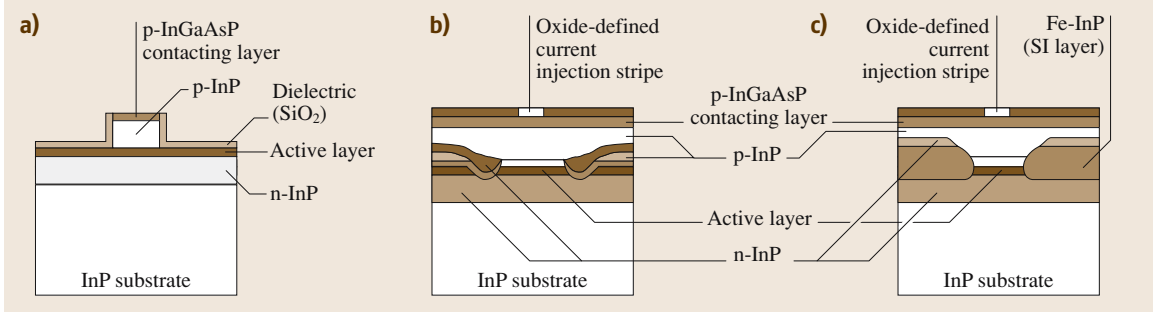


Fig. 37.12a–c Schematic diagrams of typical long-wavelength InP lasers **(a)** ridge structure, **(b)** double-channel planar buried heterostructure (DCPBH), and **(c)** semi-insulating planar buried heterostructure (SIPBH)

heterostructure (BH) devices. In double-channel planar buried heterostructure (DCPBH) lasers (Fig. 37.12b), following the growth of the layer structure, two channels, $\approx 3 \mu\text{m}$ deep and $\approx 10 \mu\text{m}$ wide, are chemically etched, forming a mesa. After cleaning, n- and p-doped current blocking layers are grown followed by the cladding and contacting layers. The current confinement is provided by the reverse-biased p–n junctions that surround the mesa.

The final structure considered is the semi-insulating planar buried heterostructure (SIPBH), as can be seen in Fig. 37.12c. In a similar way to the DCPBH devices, SIPBH lasers are based upon an etched mesa. In an SIPBH device the current confinement is provided by semi-insulating layers that surround the mesa. In these structures, after wet-chemically etching the wafer to form the mesa (typically $1\text{--}2 \mu\text{m}$ wide and $2\text{--}3 \mu\text{m}$ high) an Fe-doped layer is grown, followed by

an n-doped anti-diffusion layer, p-doped cladding and contacting layers. The resistivity of the Fe-doped regions surrounding the mesa is typically $\approx 10^8 \Omega \text{ cm}$ thereby ensuring that the current only flows through the active region. Buried heterostructure lasers tend to have lower threshold currents than ridge structures due to the combination of excellent current and optical field confinement. For high-speed applications such as telecommunications, SIPBH devices outperform DCPBH lasers due to their much lower parasitic capacitance.

37.2.11 Laser Threshold and Efficiency

In Sect. 37.2.4, the condition for material transparency was derived. In a laser structure the threshold is reached when the total gain is equal to the optical losses as the wave travels along the cavity plus the (useful) loss of light from the facets (mirror loss). If we consider the laser cavity shown in Fig. 37.8, the threshold gain condition, g_{th} is given by

$$g_{\text{th}} = \frac{1}{\Gamma} [\alpha_m + \alpha_i] \\ = \frac{1}{\Gamma} \left[\frac{1}{2L_{\text{cav}}} \ln \left(\frac{1}{R_1 R_2} \right) + \alpha_i \right], \quad (37.30)$$

where R_1 and R_2 are the facet reflectivities, α_m is the mirror loss and α_i is the internal loss per unit length in the *entire* structure including scattering losses, free-carrier absorption, and inter-valence-band absorption [37.31]. Γ is the optical confinement factor, which is equal to the fraction of the optical field intensity that overlaps the active region. This therefore accounts for the fact that not all of the optical field will give rise to gain. Note that g_{th} is the threshold gain per unit length of the device. For a typical $1000\text{-}\mu\text{m}$ -long $1.5\text{-}\mu\text{m}$ quantum well device ($\Gamma \approx 0.02$) with as-cleaved facets ($R_1 = R_2 \approx 0.3$) and $\alpha_i = 10 \text{ cm}^{-1}$, $g_{\text{th}} \approx 1100 \text{ cm}^{-1}$.

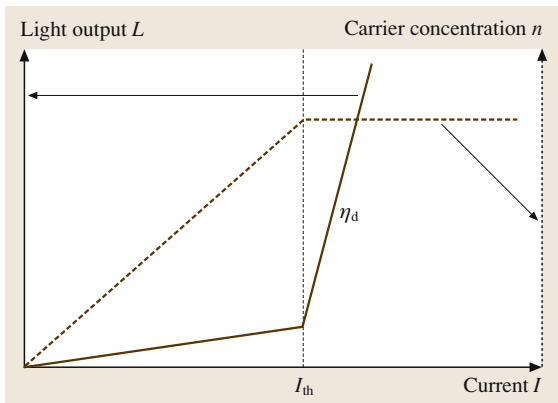


Fig. 37.13 Variation of the light output (solid curve) and carrier density (dashed curve) as a function of the injection current. Above threshold, the light output increases abruptly whilst the carrier density becomes pinned

The carrier density required to reach g_{th} is termed n_{th} , and the corresponding current (density) is termed I_{th} (J_{th}). Figure 37.13 shows how the light output and carrier density increase as current is injected into the laser structure. Below threshold, increasing the current causes the Fermi-level splitting to increase, thereby increasing the electron and hole (carrier) densities in the active region of the laser. In this regime, the laser is acting as an LED and spontaneous emission is produced. However, the electrical-to-optical conversion efficiency is relatively low ($\approx 1\%$). In practice nonradiative recombination processes (as discussed in the next section) cause the carrier density to have a nonlinear dependence on current below the laser threshold. At threshold, the stimulated lifetime decreases, causing each additional injected carrier to undergo stimulated emission. These carriers rapidly transfer to the valence band with the production of a photon via stimulated emission giving rise to an abrupt increase in the light output. Thus, at threshold the carrier density becomes pinned at n_{th} since every additional carrier quickly undergoes stimulated emission. Hence, above threshold the intrinsic *differential* quantum efficiency, η_d , of the lasing process can be as high as 100% in an ideal laser. However, effects including carrier leakage, absorption and self-heating can reduce the overall measured differential quantum efficiency.

The differential quantum efficiency, η_d , is defined as the ratio of the incremental number of photons emitted from the facets to the incremental number of carriers injected into the laser (above threshold), thus

$$\eta_d = \frac{dL}{dI} \frac{e}{h\nu}, \quad (37.31)$$

where L is the optical power emitted from both facets, e is the electronic charge and $h\nu$ is the photon energy (the lasing energy). Thus, η_d can simply be determined from the light-current characteristic (providing that all of the stimulated light is collected), η_d can more generally be defined from

$$\eta_d = \eta_i \left(\frac{\alpha_m}{\alpha_i + \alpha_m} \right); \quad (37.32)$$

η_i is the *internal* quantum efficiency and accounts for the fact that, in a real laser, not all of the injected carriers will result in the production of a photon. This may be due to inefficient injection of carriers into the active region of the laser (dependent on both the laser materials and the geometry of the laser). It is convenient

to rewrite (37.32) in terms of the inverse efficiency so that

$$\begin{aligned} \frac{1}{\eta_d} &= \frac{1}{\eta_i} \left(\frac{\alpha_i + \alpha_m}{\alpha_m} \right) = \frac{1}{\eta_i} \left(\frac{\alpha_i}{\alpha_m} + 1 \right) \\ &= \frac{1}{\eta_i} \left[\frac{2\alpha_i L_{cav}}{\ln \left(\frac{1}{R_1 R_2} \right)} + 1 \right]. \end{aligned} \quad (37.33)$$

If R_1 and R_2 are known (easily calculated), by measuring η_d for lasers with several different cavity length taken from the same wafer and plotting a graph of $1/\eta_d$ versus L_{cav} one can obtain a value for both η_i (from the intercept) and α_i (from the slope/intercept).

From Fig. 37.13 it is clear that the *overall* efficiency of the laser will depend on both the differential quantum efficiency η_d , which we wish to maximise, *and* the threshold current I_{th} , which we would like to minimise. In the discussion thus far, the threshold current has been described only in terms of a spontaneous emission current. In practice there exist many other recombination paths that contribute to the laser threshold. These are discussed in the next section.

37.2.12 Carrier Recombination Processes

When carriers are injected into a semiconductor laser there are many possible recombination paths. In an ideal semiconductor laser, above threshold, the carrier density pins, as illustrated in Fig. 37.13. Therefore, above threshold, any recombination path which depends upon n will also become pinned. However, the threshold current itself is determined by the different mechanisms by which carriers recombine in a real laser structure. In this section, the important radiative and nonradiative recombination mechanisms are briefly discussed. The threshold current of a semiconductor laser is defined as the current required to provide enough gain to reach the threshold gain level g_{th} . Due to differences between laser structures, it is frequently more useful to compare the threshold current density J_{th} between devices, which accounts for the overall area of the device that is being pumped, thereby allowing a useful comparison to be made between different device structures. In this chapter, the two are used interchangeably when discussing laser characteristics.

At threshold, the total current I can be expressed as the sum of the current paths as

$$I = eV(An + Bn^2 + Cn^3) + I_{leak} \quad (37.34)$$

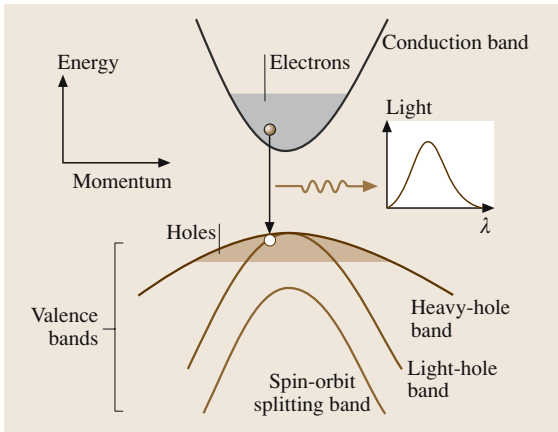


Fig. 37.14 Spontaneous emission in a bulk semiconductor

where e is the electron charge, V is the active region volume and n is the carrier density (where we assume that the electron (n) and hole (p) densities are equal, $n = p$).

The An term is due to recombination at defects and is considered to be negligible in high-quality material. It does however make significant contributions to some laser materials, most notably InGaAsN/GaAs where it can account for up to 50% of I_{th} even in the best 1.3- μm devices [37.32].

The Bn^2 term corresponds to the radiative recombination of electrons and holes giving rise to spontaneous emission, as observed from LEDs, and semiconductor lasers operated below threshold. Radiative recombination occurs (for interband processes) when an electron in the conduction band recombines with a hole in the valence band, resulting in the spontaneous emission of a photon as illustrated in Fig. 37.14. For lasing to occur it is necessary to have spontaneous recombination to initiate the stimulated emission process. However, because spontaneous emission is a random process, emitting photons in all directions over a broad energy range, only a small fraction of the photons can couple to the laser gain. Hence for efficient laser operation it is desirable that the amount of unwanted spontaneous emission is reduced.

The Cn^3 term describes *nonradiative* Auger recombination. In an Auger recombination process, the energy of a recombining electron-hole pair that would normally produce a photon, is instead given to a third carrier (electron or hole) which is excited further into its respective band. Thus, in an Auger recombination process, three carriers are involved, the initial electron-hole pair plus an additional electron or hole.

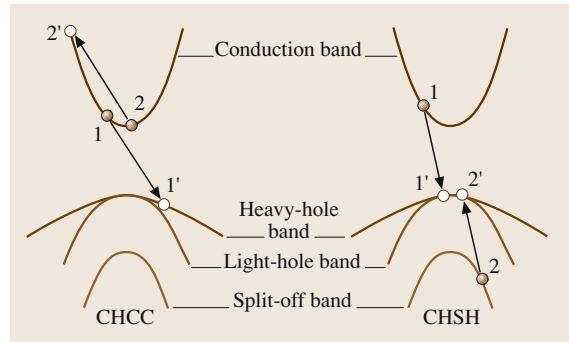


Fig. 37.15a,b The two primary Auger recombination processes: (a) CHCC producing hot electrons and (b) CHSH producing hot holes

The two main Auger recombination processes are illustrated in Fig. 37.15, where the closed circles represent electrons and open circles represent holes. In the conduction-hole-conduction-conduction (CHCC) process, the energy and momentum of a recombining conduction-band electron and valence-band hole are transferred directly to a second conduction-band electron, which is then excited into a higher-energy conduction-band state. In the direct conduction-hole-spin-hole (CHSH) process, the energy and momentum are transferred to a second valence-band hole, which is then excited into the spin split-off band. In addition to these direct Auger recombination processes, there may be phonon-assisted processes whereby the conservation

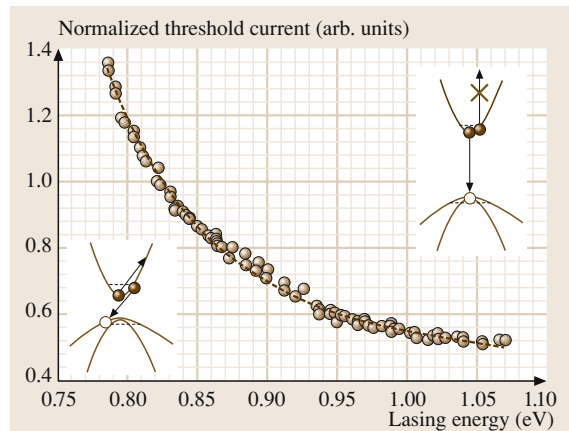


Fig. 37.16 Variation of the threshold current with lasing energy over the wavelength range 1.2–1.6 μm . With increasing wavelength (decreasing energy), the threshold current increases swiftly due to the increased probability of Auger recombination, as illustrated in the insets

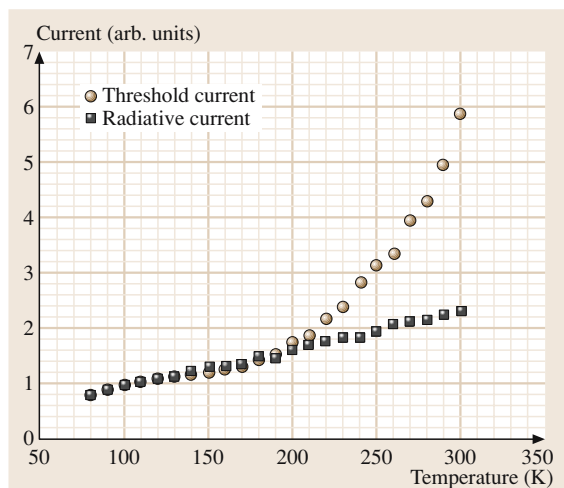


Fig. 37.17 Threshold current (circles) and radiative current (squares) versus temperature for a 1.3- μm InGaAsP laser. The threshold current increases more strongly with temperature than the radiative current due to the thermally induced onset of Auger recombination

of momentum and energy of the carriers is relaxed by the emission or absorption of a phonon. These are however of less importance in 1.3- μm and 1.5- μm lasers [37.33]. It should be noted that Auger-generated carriers are very energetic and initially lie many $k_B T$ more than E_g above their respective band edges. Whilst these carriers quickly scatter back to thermal equilibrium, there is a finite probability that they will scatter from the active region into the cladding layers, thus forming an additional Auger-generated leakage current.

Auger recombination processes are very sensitive to the band gap, E_g , and are severely detrimental to laser performance. For emission wavelengths below 1 μm , the probability of an Auger recombination process occurring is small due to the near-vertical transitions that are required on an energy-momentum diagram, which decreases the probability of the process occurring (inset Fig. 37.16). However, as E_g decreases, the transitions become less vertical and the probability of finding a state for the *hot* Auger carrier increases. Auger recombination therefore begins to influence the performance of semiconductor lasers and by 1.5 μm it dominates their behaviour at room temperature, forming $\approx 80\%$ of the threshold current [37.28]. This is illustrated in Fig. 37.16 where the normalised variation of the measured threshold current with lasing energy was determined by performing high-pressure measurements on several lasers over the nominal wavelength

range 1.3–1.6 μm . With increasing lasing energy (band gap), the threshold current decreases strongly due to the lowered contribution of Auger recombination. Thus, long-wavelength lasers are more susceptible to Auger recombination. Auger recombination is also largely responsible for the high-temperature sensitivity of the threshold current of 1.3- μm and 1.5- μm semiconductor lasers. In Fig. 37.17 we plot the temperature dependence of the threshold current (circles) for a 1.3- μm InGaAsP laser. The squares correspond to the measured radiative current as determined from spontaneous emission measurements [37.28]. At low temperature, the threshold current follows the radiative current very closely, however, above ≈ 160 K, the threshold current increases strongly due to thermally activated Auger recombination. This is primarily due to the n^3 dependence of the Auger current but also because the Auger coefficient C is thermally activated [37.34].

I_{leak} is associated with heterobarrier leakage and current spreading around the active region. With increasing temperature there is an increasing probability that electrons in the conduction band and holes in the valence band will have enough thermal energy to escape from the quantum wells. This is termed heterobarrier leakage. Here these carriers may recombine radiatively (at shorter wavelengths) or nonradiatively. If they have sufficient energy, they may even escape into the cladding layers. The probability of carriers escaping into the cladding layers depends upon the position of the quasi-Fermi levels with respect to the cladding band edge, as illustrated in Fig. 37.18. With increasing temperature, the thermal spread of carriers broadens and the population of unconfined carriers at threshold increases, thereby increasing the threshold current. For diffusion-driven leakage of electrons, the leakage current I_{leak} can be described via an activation energy E_a , equal to the difference in en-

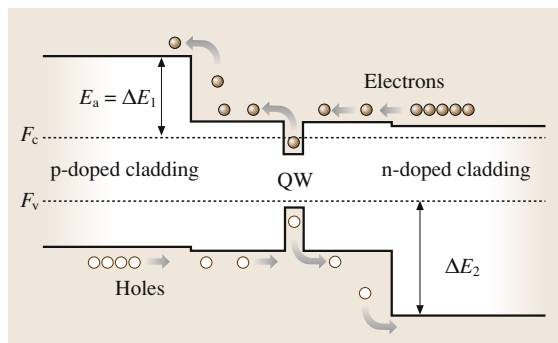


Fig. 37.18 Carrier leakage in a quantum well (QW) semiconductor laser

ergy between F_c and the lowest (Γ , L or X) CB edge of the cladding layer. Thus

$$I_{\text{leak}} = I_0 \exp\left(-\frac{E_a}{k_B T}\right), \quad (37.35)$$

where k_B is the Boltzmann constant, T is the temperature and I_0 is independent of temperature. As with Auger recombination, carrier leakage is strongly temperature-dependent and can lead to a large change in the threshold current with temperature. For this reason, much effort has been put into designing lasers which minimise carrier leakage. In III–V semiconductors, due to the large hole effective mass, F_v moves towards the valence band at a much slower rate than F_c moves towards the conduction-band edge. For InGaAs(P)-based materials as used in 1.3- μm and 1.5- μm lasers, the quantum well to barrier/SCH band offset in the conduction band is much lower than the valence-band offset. Additionally, holes have a much lower mobility than electrons. Consequently, the overflow of electrons is generally considered to be more significant compared to the overflow of holes in this materials systems. Carrier leakage is a particular problem for visible semiconductor lasers for which the Al-containing layers may be close to, or indeed are, indirect, whereby the conduction-band X-minima are at a lower energy than the conduction band Γ -minimum. Thus, electrons may thermally escape from the Separate confinement Heterostructure (SCH) into the cladding, forming a leakage current. An example of this is shown in Fig. 37.19 where the threshold current is plotted as a function of temperature for a 670-nm Al(GaInP) laser.

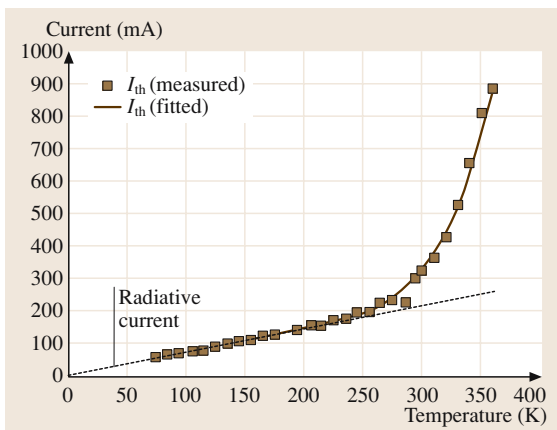


Fig. 37.19 Threshold current versus temperature for a 670-nm GaInP quantum well (QW) laser. The strong temperature sensitivity above 250 K is due to electron leakage into the X-minima of the AlGaInP barrier layers

The strong increase in threshold current with temperature is due to the thermal leakage of electrons from the GaInP QW via the SCH region into the X-minima of the AlGaInP barrier layers. At room temperature, leakage accounts for $\approx 20\%$ I_{th} at room temperature, rising to $\approx 70\%$ at 80°C at this wavelength [37.35]. Due to the smaller activation energy, shorter-wavelength devices suffer still further from carrier leakage and this, in practice, sets the lower wavelength limit for lasers produced using this material system.

37.2.13 Temperature Sensitivity and T_0

As described in the previous section, the threshold current of lasers operating in the visible and in the 1.3- μm to 1.5- μm range is very sensitive to the ambient temperature. From a commercial perspective, this is very unsatisfactory as these lasers are required to provide a constant light output for a given drive current above threshold all year round. However, throughout the year, the ambient temperature can typically vary significantly. Hence, further cost results from having to use expensive temperature-control electronics to stabilise the temperature and hence light output of these lasers. This can be seen in Fig. 37.20 where the power–current curves are plotted for a 1.3- μm laser over the temperature range 20 – 70°C . Clearly, by operating the laser at a constant bias current, the light output decreases considerably as the temperature decreases. In extreme cases the current may no longer be sufficient to reach laser threshold, at which point the device will no longer lase and the output would only be spontaneous emission (as in Fig. 37.20,

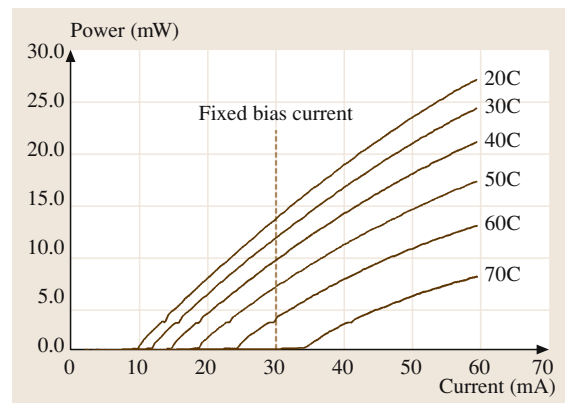


Fig. 37.20 Typical curves of power versus current for a 1.3- μm laser as a function of temperature. The large increase in threshold current with increasing temperature gives rise to a strong loss of power for a fixed bias current

for the 70 °C curve). In practical situations this therefore necessitates the use of temperature-control feedback electronics. These add considerable cost (many times the cost of the laser itself) to the system. This has become known as the T_0 problem. Around room temperature and over a limited range of temperature, the threshold current of many semiconductor lasers was observed to increase approximately exponentially with temperature [37.36]. Hence it is usual to write that $I_{th} = I_0 \exp(T/T_0)$, where I_0 is a constant and T_0 is the characteristic temperature of the threshold current given by

$$\frac{1}{T_0} = \frac{1}{I_{th}} \frac{dI_{th}}{dT} = \frac{d \ln(I_{th})}{dT}, \quad (37.36)$$

T_0 is usually expressed in units of Kelvin. It follows that a low value of T_0 corresponds to a high temperature sensitivity and vice versa. Hence, a high T_0 is desirable. Experimentally, T_0 can be determined from measurements of I_{th} from

$$T_0 = \frac{T_2 - T_1}{\ln \left[\frac{I_{th}(T_2)}{I_{th}(T_1)} \right]}, \quad (37.37)$$

where $I_{th}(T_1)$ and $I_{th}(T_2)$ are the threshold currents at T_1 and T_2 respectively and $T_2 > T_1$. In practice, over a wider temperature range, T_0 is *itself* temperature-dependent due to the temperature dependence of the recombination mechanisms that occur in the lasers. However, T_0 is a parameter often used to investigate the temperature dependence of the threshold current. For a typical 1.3- μm or 1.5- μm QW laser, $T_0 \approx 50$ –60 K. This is due to Auger recombination, as discussed above. It can be shown that, for an ideal QW laser, the maximum $T_0 = 300$ K around room temperature, where the threshold current is entirely due to radiative recombination [37.37]. Furthermore, by going to quantum dot lasers, due to the 3D-confinement effect, an infinite T_0 has been predicted [37.12]. Recently reported experimental work has shown that this may be possible by using p-doped quantum dot layers, which exhibit temperature-insensitive operation around room temperature [37.38].

37.3 Single-Mode Lasers

37.3.1 DFB lasers

In an ideal laser, the output spectrum would be very narrow with no side modes. In practice, the output spectrum of Fabry–Perot devices such as those described earlier in this chapter consists of several competing modes. This is due to the comb of Fabry–Perot modes, which overlap with the laser gain spectrum (shown earlier in Fig. 37.9). There are a number of laser devices which can be operated with high spectral purity. Perhaps the most important of these is the distributed feedback (DFB) laser. In a Fabry–Perot laser, the end facets provide optical feedback to give rise to provide sufficient gain to achieve threshold. Because the reflectivity of the end facets is only slightly wavelength-dependent (for

a given polarisation), this gives rise to several longitudinal modes. These are clearly visible in the output amplified spontaneous-emission spectrum of a Fabry–Perot laser, indeed this is one way in which the gain can be measured [37.39]. In contrast, in a DFB laser, there is an additional layer called the guiding layer (Fig. 37.21) grown close to the active layer. The guiding layer is essentially a grating consisting of a corrugated layer of a dissimilar-refractive-index semiconductor defined using lithographic and etching techniques. The guiding layer is usually overgrown although surface gratings are also possible and simpler to produce. The corrugations in the material provide a periodic variation in the refractive index along the laser cavity and hence the optical feedback is *distributed* along the length of the cavity. The forward and backward propagating waves will only interfere constructively if their frequency is related to the pitch of the grating. It can be shown that the correspondingly allowed modes lie symmetrically either side of the Bragg frequency of the grating. In practice, due to nonuniformity as a result of processing, or by artificially introducing a quarter-wavelength shift in the Bragg reflector, one of the modes may lase preferentially. DFB lasers are the mainstay of optical communications systems due to their single-mode behaviour with

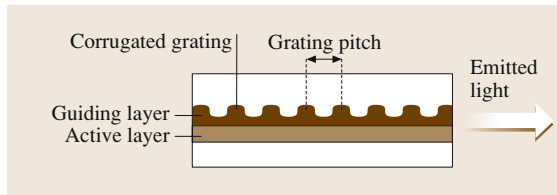


Fig. 37.21 Simplified diagram of a distributed feedback (DFB) laser

high side-mode suppression and narrow line widths (≈ 0.1 nm). In an optical-fibre communication system, this allows for the simultaneous transmission of several data channels with closely spaced wavelengths around the fibre attenuation minimum of $1.55\ \mu\text{m}$. this is known as dense wavelength-division multiplexing (DWDM).

An added benefit of the grating in a DFB laser is that it may be used to artificially reduce the temperature sensitivity of the laser threshold current and lasing wavelength. Unlike Fabry–Perot lasers, which primarily emit at the mode closest to the gain peak, as discussed above DFB lasers can be made to emit at one fixed mode. The peak of the gain spectrum moves approximately at the same rate as that of the band gap ($\approx 0.5\ \text{nm/K}$) whilst the temperature dependence of the DFB mode follows the thermally induced change in the refractive index ($\approx 0.1\ \text{nm/K}$) which is much less temperature sensitive. Thus, DFB lasers are far less susceptible to thermally-induced mode hops. Furthermore, by carefully designing the peak of the gain spectrum to lie at a shorter wavelength than the DFB wavelength, one can effectively *tune* the DFB laser with increasing temperature resulting in, over a limited temperature range, a relatively temperature-independent threshold current. This has been demonstrated as a means of achieving cooler-less $1.3\text{-}\mu\text{m}$ lasers, albeit at the expensive of higher threshold currents.

37.3.2 VCSELs

Vertical-cavity surface-emitting lasers (VCSELs) differ from conventional semiconductor lasers due to the fact that the optical cavity lies in the growth direction, in the same direction as the current flow. The output of the laser is therefore emitted from the surface. There are several major advantages of VCSELs compared with edge-emitting lasers, which have led to a considerable amount of research being undertaken to produce VCSELs at a variety of wavelengths. Unlike edge-emitting lasers, once the semiconductor wafer has been metallised, it is very easy to test the devices prior to dicing them into individual chips. This represents a considerable cost saving. Forming laser arrays is obviously therefore much more straightforward with VCSELs and has led to the development of parallel data links based upon VCSELs and high-power VCSEL arrays. The small size of VCSELs (aperture diameter $< 10\ \mu\text{m}$) also gives rise to both single-mode behaviour and low threshold currents. Furthermore, due to the fact that the light is (usually) emitted from a circular geometry contact, the beam profile is circular, making fibre coupling much

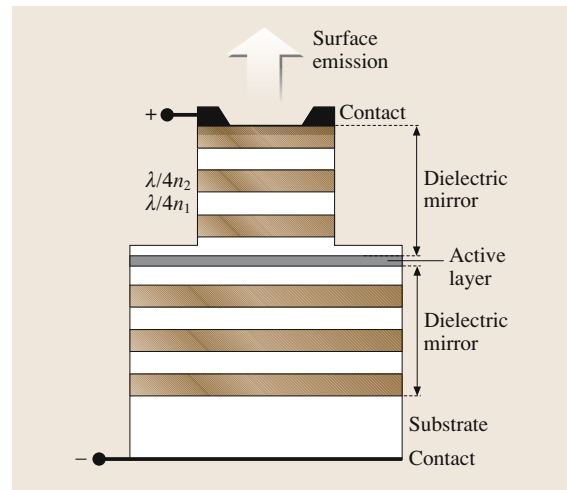


Fig. 37.22 Schematic diagram of a vertical-cavity surface-emitting laser (VCSEL)

simpler. In contrast, edge-emitting lasers have an elliptical beam profile due to diffraction of the laser light from the end facet, requiring expensive corrective optics and active alignment with optical fibres.

Because the VCSEL emits from its surface, the optical cavity over which the optical field may experience gain is very short. Consequently, in order to reach the threshold gain level, the optical field needs to make several passes through the active region. This requires highly reflective end mirrors. At a typical III–V semiconductor–air interface, the refractive-index contrast gives rise to a reflectivity of $\approx 30\%$. This is enough to achieve sufficient gain to reach threshold in an edge-emitting laser, which is typically many hundreds of microns long. However, in a VCSEL, the cavity length is typically $\approx 0.1\ \mu\text{m}$, requiring mirror reflectivities $> 99\%$. This is achieved by sandwiching the active region between two multilayer dielectric mirrors (Fig. 37.22). These so called distributed Bragg reflectors (DBRs) consist of alternating high- and low-refractive-index quarter-wave-thick pairs. In a similar way to the DFB lasers discussed above, the DBR mirrors in VCSELs provide a high degree of wavelength-selective reflectance, providing that the sum of the optical thicknesses of each layer pair is equal to a half wavelength (i.e. $n_1d_1 + n_2d_2 = \lambda/2$, where n_i and d_i are the refractive index and thickness of each layer respectively), thereby giving rise to constructive interference at the desired wavelength. Typically, there may be 30 or so layers in each DBR to achieve sufficiently high reflectivity. It is useful to note here that a related device, the

resonant-cavity LED (RCLED), is similar to the VCSEL but contains far fewer DBR pairs in order that it does not lase, but instead produces narrow-band high-efficiency spontaneous emission. Whilst a large number of layer pairs in the VCSEL gives rise to a high reflectivity it may also decrease the quantum efficiency of the device and increase the operating voltage due to the large number of interfaces. Furthermore, due to free-carrier absorption and carrier leakage in the DBRs themselves, a high number of DBRs can degrade device performance [37.40]. The exact number of DBRs used is therefore a compromise which depends upon the operating wavelength of the VCSEL and the semiconductor material system on which it is based.

High-quality VCSELs have already been produced for emission at 650 nm and 850 nm (for plastic and silica fibre communications), for oxygen sensing using 760 nm VCSELs and high-power devices for emission at 980 nm [37.41]. However, there has been considerable difficulty in producing VCSELs at the telecommunications wavelengths of 1.3 μm and 1.55 μm . This is primarily due to the fact that GaAs/AlGaAs is the preferred system for making highly reflective DBRs due to the high refractive-index contrast between GaAs and AlAs. Such DBRs are therefore compatible with the GaAs substrates as used in devices below $\approx 1 \mu\text{m}$. However, for longer-wavelength devices, which are primarily based on InP substrates it is difficult to form highly reflective DBRs using InGaAsP/InP due to their low refractive-index contrast. Alternative methods of producing long-wavelength VCSELs are inherently process-intensive, using techniques such as wafer fusion in which the active layer and InGaAsP/InP DBR is first grown on an InP substrate whilst a GaAs/AlAs DBR is grown separately on a GaAs substrate. The active layer is then fused onto the GaAs/AlAs DBR and the InP substrate is then removed (see for example, [37.42]). The wafer fusion process is understandably both costly and difficult, and therefore methods of growing monolithic 1.3- μm and 1.55- μm VCSEL structures

using only a single GaAs substrate are particularly attractive. This has stimulated a great deal of research into producing GaAs-based active regions emitting at these wavelengths. In recent years there has been some success in achieving this by two approaches. The first approach is to use InAs/GaAs quantum dots as the active region. Quantum dot lasers with very low threshold current density ($< 20 \text{ A/cm}^2$) emitting at 1.3 μm have already been realised [37.13]. However, there are few reports in the literature on quantum-dot-based VCSELs. This is largely due to the difficulty in achieving sufficient gain from the quantum dots and can result in the need for several layers of quantum dots. In spite of this, quantum dots are showing promise for use in temperature-insensitive edge-emitting lasers and high-power lasers.

The second approach has been in the use of the so called dilute nitrides. In 1997, *Kondow* and co-workers originally proposed the use of InGaAsN/GaAs as an active material to achieve long-wavelength emission on GaAs [37.43]. This is due to the unusual band-gap bowing that occurs when small concentrations ($\approx 2\text{--}5\%$) of nitrogen are added to GaAs. By growing the alloy InGaAsN, one may achieve long-wavelength emission whilst maintaining a low or zero strain. Although material quality remains an issue, there have been several successful demonstrations of edge-emitting lasers and VCSELs based upon this material emitting at 1.3 μm [37.44–46]. The push towards 1.55- μm emission has been largely hampered by material quality issues although at the time of writing this chapter, there have been the first reports of low-threshold 1.5- μm edge-emitting lasers based upon InGaAsN/GaAs [37.47]. Other researchers have produced close-to-1.5- μm edge-emitting and VCSEL devices with the pentenary material InGaAsNSb/GaAs [37.48]. It is speculated that antimony may act as a surfactant to improve the quality of growth. The dilute nitride approach does at present appear to offer the best possibility of obtaining VCSELs emitting at the technologically important wavelength of 1.55 μm .

37.4 Optical Amplifiers

37.4.1 An Introduction to Optical Amplification

As a light pulse propagates through an optical communication system, the pulse becomes attenuated until eventually it is necessary to regenerate the pulse to keep

the signal above the background noise level. This was initially achieved by detecting the pulse using a photodetector, and then using the detected electronic pulse to trigger a laser giving a fresh output pulse. An alternative technique is to use direct optical amplification. This has been achieved for instance by doping optical fi-

bres with rare-earth elements such as erbium, forming an erbium-doped fibre amplifier (EDFA). Population inversion is achieved by exciting the erbium atoms at 980 nm or 1.48 μm , where the excited carriers decay to a level from which they can recombine to give stimulated emission at 1.55 μm , amplifying weak signals in the fibre. Amplification can also be achieved within conventional silica fibres using the Raman effect; by pumping the fibre with a high-power pump laser (total power $\approx 1\text{ W}$) at a fixed frequency above the signal frequency, Raman scattering gives rise to gain at a lower frequency (the difference is the phonon energy). Thus, by pumping an optical fibre at 1.45 μm one may achieve gain at 1.55 μm . The main advantage of Raman amplification compared with EDFAs is the fact that the gain curve may be dynamically tuned by judiciously using different wavelength pump lasers. The disadvantages of Raman amplification is that the overall gain is generally lower than an EDFA and cross-amplification effects can be a problem. For some applications, semiconductor optical amplifiers (SOAs) may be preferred.

37.4.2 Semiconductor Optical Amplifiers (SOAs)

The structure of a SOA is very similar to that of a laser, but with one significant difference: the reflectivity of the end facets, $R \approx 0$. This is achieved through the use of multilayer antireflection coatings together with angled facets or cavities. This suppresses lasing within the cavity, and eliminates reflected signals in the optical system.

The signal emerging from an optical fibre is, in general, randomly polarised and therefore it is very desirable that any optical amplifier provides gain that is independent of the direction of polarisation of the light. In a bulk heterostructure device, the TE and TM material gain are equal. However, the optical confinement factor Γ for

TE light polarised in the plane of the heterostructure is slightly larger than for TM light polarised perpendicular to the heterostructure plane, leading to a larger overall TE gain. As discussed earlier in this chapter, in a normal quantum well system the quantum confinement brings the heavy-hole band to the top of the valence band and light polarised in the TE mode is amplified considerably more strongly than light polarised in the TM mode. This problem can be overcome using strained-layer techniques. It is possible to grow the wells or the barriers with a small amount of tensile strain. This raises the light-hole band and the strain can be adjusted to increase the device TM gain until it is just equal to that of the TE gain. Work to date indicates that the relative gains in the two modes are sensitive to the magnitude of the amplifier current and so the gains are equal over a limited range.

Strained SOAs

Another more promising approach is to introduce both compressive- and tensile-strained wells alternately into the active region. The compressive wells provide gain predominantly to the TE mode and the tensile wells predominantly to the TM mode. The structure can then be designed to make the two gain modes equal. The initial work by Tiemeijer et al. from Philips concentrated on 1.3- μm devices [37.49]. They found that a combination of four compressive wells and three tensile wells each with 1% strain gave TE and TM gain within 1 dB of each other over a wide wavelength band and over an order of magnitude change in amplifier current. This structure is an excellent example of the flexibility in device design that is afforded by the introduction of strained-layer techniques. Other applications of such a structure include two-polarisation or two-frequency lasers, polarisation control elements and nonabsorbing strain-overcompensated mirrors in high-power lasers.

37.5 Modulators

37.5.1 Modulator Theory

Although it is possible to modulate the output from a semiconductor laser by directly modulating the injected current, in some applications requiring high-speed or low-power switching, it is advantageous to run the laser source CW and modulate the light subsequently [37.50]. This is basically because, unlike lasers,

semiconductor modulators are diodes operated in reverse bias. They therefore draw little current and, because there is no carrier injection, there is much less chirp.

Electroabsorption modulators make use of the fact that the presence of a large electric field adds a perturbation to the Schrödinger equation, which leads to a decrease in the band-gap energy E_g . Thus photons with

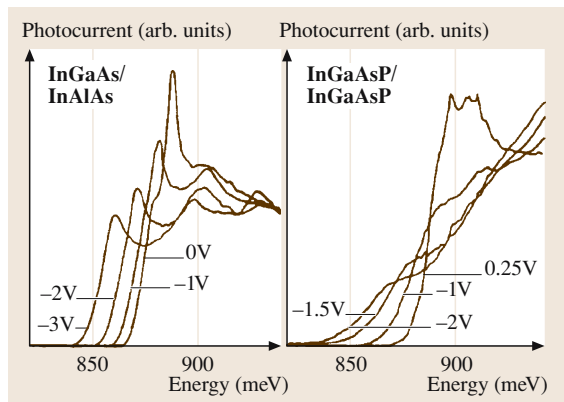


Fig. 37.23 Measured photocurrent spectra at 2 K of an InGaAs/InAlAs (*left*) and an InGaAsP/InGaAsP (*right*) electroabsorption modulator at several bias voltages. (After [37.50] with permission)

energies just below the band gap become more strongly absorbed when a field is applied, leading to a decrease in light intensity. Electroabsorption modulators can benefit enormously from the use of multiple quantum wells, where the electroabsorption effect is more than one order of magnitude greater than that in comparable bulk structures.

In quantum wells the modulation of the energy levels by an electric field applied perpendicular to the plane of the wells is called the quantum-confined Stark effect (QCSE). For a given applied voltage, it causes more change, $\Delta\alpha$, to the absorption coefficient α close to the band edge than in equivalent bulk structures for three main reasons. Due to the quantum confinement, there is a step-like increase in the density of states at the band edges, as described above, and so a faster rate of increase in absorption coefficient with energy. Thus, $\Delta\alpha/\Delta E_g$, the change in absorption coefficient with band gap, is larger. Secondly, the steep rise in the absorption coefficient with energy at the band edge is further enhanced in quantum well structures due to excitonic effects. An exciton consists of an electron–hole pair bound together by electrostatic attraction and has associated with it a sharp line absorption spectrum which is superimposed on the quantum well absorption. Figure 37.23 shows the measured photocurrent spectra at low temperature as a function of the applied reversed bias for two long-wavelength modulator structures. The exciton peak is clearly visible at the absorption edge, particularly for low bias values. Excitons also exist in bulk material at low temperatures but not at room temperature. In quantum wells however, the electrons and holes are confined

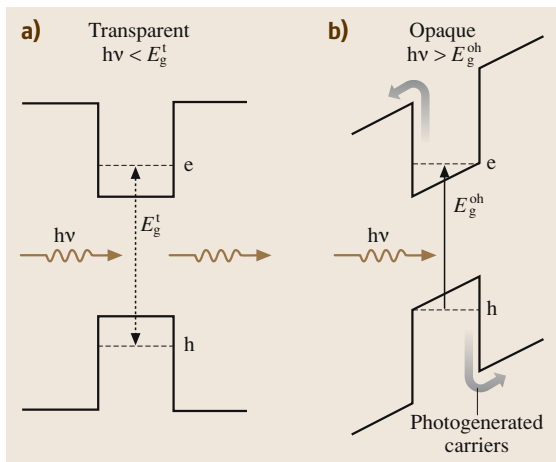


Fig. 37.24a,b The band alignment of a QW in (a) flat band (zero field) where the optical gap equals $E_g^t > h\nu$ (transparent) and (b) for a large field across the QW, where the optical gap equals $E_g^0 < h\nu$ (opaque)

into the same region of space by the well and so the electrostatic binding energy is sufficiently large that they are still able to exist at room temperature. Finally, the actual energy of the band edge is largely determined by the effect of quantum confinement, which depends not only on the width of the quantum well but also on its shape. Since the shape of the well can be strongly modified by an applied electric field, so can the electronic energies at the band edges, as can be seen from the shift in absorption edge with applied bias in Fig. 37.23. Thus ΔE_g is enhanced by quantum confinement.

The effect of an electric field applied perpendicular to the plane of a quantum well is to change the shape of the well from the rectangular well shown in Fig. 37.24a to the triangular shape shown in Fig. 37.24b. Solutions to the Schrödinger equation show that the lowest confined states are closer together in the triangular situation than the rectangular one and so the effective absorption edge is moved to lower energies. Because the electron ground state is shifted towards one side of the quantum well, and the hole ground state towards the opposite side, the overlap of the electron and hole wavefunctions is decreased by the application of an electric field. The result is to decrease the absolute value of the absorption coefficient and to decrease the exciton absorption peak, as seen in Fig. 37.23. However, if one considers an energy just below the band-edge exciton energy at zero field, at about 850 meV in Fig. 37.23, the absorption coefficient at that energy can be greatly increased by the application of an electric field.

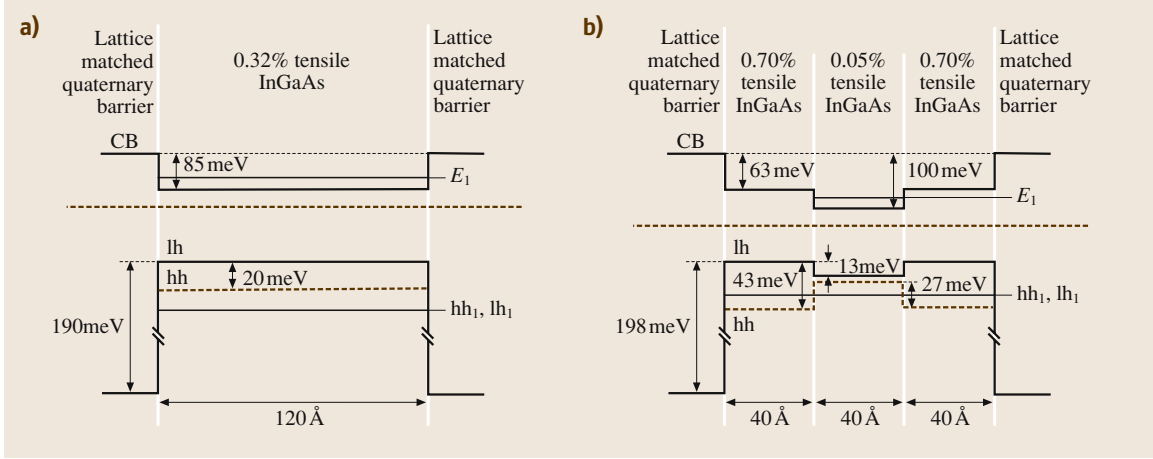


Fig. 37.25a,b Calculated conduction- and valence-band profiles for (a) conventional tensile QW and (b) multi-strain-stepped QW (After [37.51])

For a modulator, a very important parameter is the ratio of the light intensity transmitted in the *on* state to that transmitted in the *off* state when the field is applied. This is given simply by

$$P = \exp(\Delta\alpha L), \quad (37.38)$$

where L is the distance the light travels through the absorbing medium. In quantum well structures, although $\Delta\alpha$ is relatively large, if the light travels in a direction perpendicular to the plane of the wells, the length L is small; of the order of tens of nm. One way to overcome this problem is to use the QCSE modulator in the waveguide configuration; i. e. to pass the light parallel to the plane of the quantum wells in a waveguide structure similar to that for a laser as described previously. In this case the effective absorption per unit length will be decreased by the optical confinement factor (Γ) to $\Gamma\alpha$, but L can easily be increased much more; to tens or even hundreds of μm . Such a waveguide configuration can achieve a large on/off ratio but it has two potential problems. Firstly, it can have a large insertion loss because it is difficult to launch the light into the waveguide unless it is integrated with the laser source and also there can be light scattering and absorption along the length of the guide even in the on state. Secondly, the modulator will be dependent on the polarisation of the light beam travelling along the waveguide, as described below.

37.5.2 Polarisation-Insensitive Modulators

As was explained when considering strained-layer electronic properties, the degeneracy of the heavy- and

light-hole bands, which exists in bulk material, is split by confinement effects in a quantum well. This is because the quantum confinement energy is proportional to l/m^* and so the first light-hole valence band lh_1 lies below the first heavy-hole valence band hh_1 . As a consequence, the wavefunctions that make up the edge of the valence band are derived from the heavy-hole band and so only interact with TE-polarised photons as described above. Thus, while photons of energy close to the band edge and polarised in the TE mode are modulated as the band edges move together with applied electric field, photons of the same energy but polarised in the TM mode, are relatively unaffected. This is in general unacceptable if the photons are arriving from a system, such as an optical fibre, which may deliver photons of any polarisation in a manner which is sensitive to extraneous influences.

The problem of polarisation sensitivity in modulators, as with the polarisation-insensitive amplifier, may be dealt with by engineering the band structure with a judicious use of in-built strain within the quantum wells. As described above, the light-hole band may be raised in energy with respect to the heavy-hole band by the introduction of tensile strain. Thus, with tensile strain it is possible to shift the heavy- and light-hole bands until the confined states are once again degenerate, as shown in Fig. 37.25a. Such a system, although it may be polarisation insensitive at zero net field, can again become polarisation sensitive as a net field is applied, as is shown by the movement of the band edges in Fig. 37.26a. This effect arises since the heavy-hole confined level, being closer to its band edge, is more sensitive to the distortion of that band edge by the applied field. Once again, how-

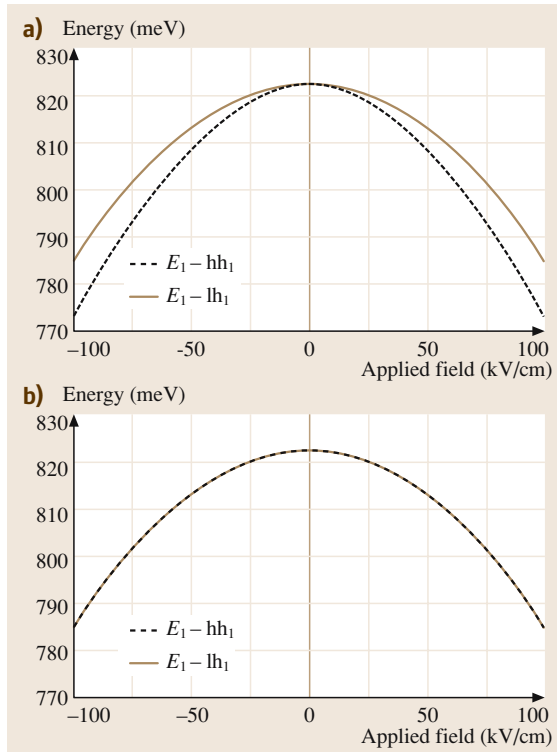


Fig. 37.26a,b Calculated quantum confined Stark shifts of the $E_1 - hh_1$ (dashed lines) and $E_1 - lh_1$ (solid lines) transitions for structures (a) and (b) in Fig. 37.25. (After [37.51])

ever, it is predicted that this effect can be compensated by further band-structure refinements brought about by employing three zones of different strain within the same well. This is illustrated in Fig. 37.25b. In this multi-strain well, the tensile strain is larger in the outer thirds of the well. Thus, when a field is applied and the hole wavefunctions are moved towards the edge of the well, the heavy-hole level sees a smaller effective well width and its change in energy can be made the same as for the light hole. Figure 37.26b shows that excellent matching of the movements of the heavy- and light-hole bands can be achieved with the three-zone well over the full range of practical electric fields, up to 100 kV/cm.

37.5.3 High-Speed High-Power QCSE Modulators

Strained-layer QCSE modulators have been successfully integrated with strained-layer lasers to produce high-speed low-chirp sources operating at 1.55 μm for optical communications. The laser and the modulator

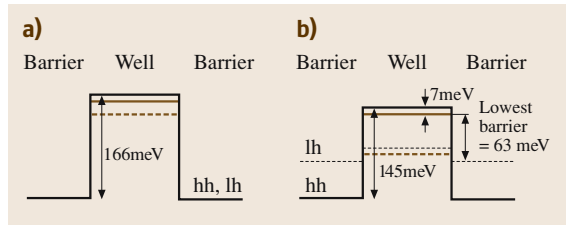


Fig. 37.27a,b Calculated valence-band edge energies and band alignments for (a) lattice-matched modulator and (b) strained modulator, as in Fig. 37.28. The thick horizontal lines indicate the highest hh confined state

consist of the same wave-guide structure but are used in forward and reverse bias respectively. In this case there are no polarisation problems because if, for example, compressively strained wells are used the laser will produce the TE-polarised output, which is efficiently modulated by the modulator. One problem that is encountered, however, is that while the modulator is in the absorbing state a significant density of photogenerated carriers may accumulate in the quantum wells. This has two effects. Firstly, the carriers tend to fill the band-edge states, thus reducing the field in the well, and in addition bleaching the absorption process. Secondly, they may take a long time to escape, so that efficiency of switching to the nonabsorbing state is dependent on how long the modulator was absorbing and what carrier density had accumulated. In practice it is found that the escape rate of photogenerated holes from the quantum wells is slower than that for electrons and is the limiting process. This problem may be addressed by changing from the standard InGaAsP material system to InGaAlAs, which has a much smaller valence-band offset than in the equivalent InGaAsP layers [37.50]. Alternatively, the problem may be reduced by band-structure engineering in InGaAsP using strain.

Figure 37.27a shows the valence-band structure of a lattice-matched InGaAsP-based electroabsorption modulator designed to operate around 1.5 μm . The heavy-hole well was calculated to be 166 meV deep, and the structure showed power saturation, as illustrated by the open squares in Fig. 37.28. However, when a 1% compressive strain was introduced into the well and 1% tensile strain into the barriers between the wells, the calculated excitation energy from the heavy-hole states in the well to light-hole states in the barrier was reduced to 63 meV, as illustrated in Fig. 37.27b. Since the strain in the barrier has little effect on the well depth for heavy holes, the heavy-hole wavefunctions and quan-

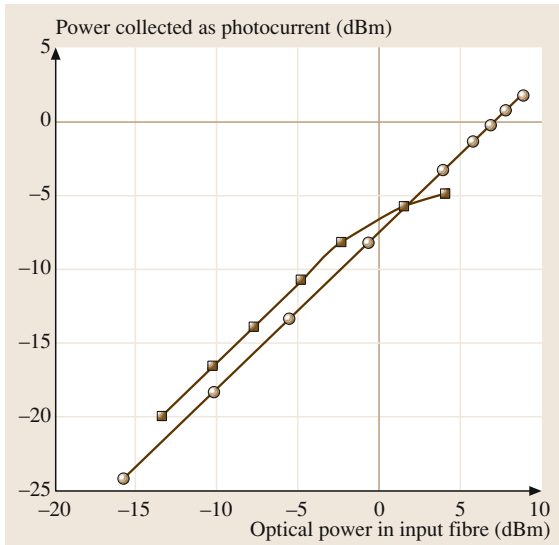


Fig. 37.28 Total power absorbed in lattice-matched (*open squares*) and strained (*closed squares*) electroabsorption modulator as a function of power in the input fibre. (After [37.52] with permission)

tum confined Stark effect should be comparable in both structures, but now thermal excitation out of the well by phonon scattering to the light-hole level in the barrier is greatly enhanced. The improvement in the power-saturation characteristics of the strained modulator are shown by the solid squares in Fig. 37.28; a considerable improvement in the 10-GHz switching performance was also observed.

37.5.4 The Electro-Optic Effect

The real part of the refractive index can be described in terms of virtual transitions between the valence and conduction band and is therefore related to the imaginary part of the refractive index. Hence, when an electric field applied to a semiconductor leads to a change in the absorption characteristics as described above, there is also

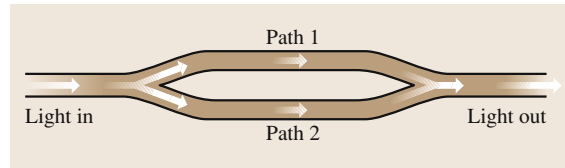


Fig. 37.29 Schematic of a Mach-Zender interferometer

an accompanying change in the real part of the refractive index. This results in a phase change of the light reaching the end of the active region of the modulator. Whilst this causes unwanted chirp in the electroabsorption modulators described above, it can be used to advantage in modulators and switches that employ interference effects. A good example of such a device is the modulator based on a Mach-Zender interferometer. In this structure the incoming wave is divided into two equal components at a Y-junction and then recombined at a similar junction, as shown in Fig. 37.29. If the two arms of the interferometer are exactly equivalent then the two beams will arrive in-phase and constructively recombine at the output. However, if the refractive index for light travelling down one path is increased by the electro-optic effect so that its phase at the output is changed by π , then the two beams will destructively interfere at the output and no light will be transmitted. Band-structure considerations for the electro-optic effect are ones with which we are already familiar. The line-width enhancement factor α_1 is defined by

$$\alpha_1 = -\frac{4\pi}{\lambda} \left(\frac{\frac{dN}{dg}}{\frac{dg}{dn}} \right) \quad (37.39)$$

and provides a measure of the ratio of the change in the real part N of the refractive index to the change in the imaginary part g of the refractive index, each as a function of carrier density n . Thus, for modulators based on the electro-optic effect, we would wish to have a band structure that results in a large α_1 , while for electroabsorption modulators we require α_1 to be as small as possible.

37.6 Photodetectors

In the first part of this chapter we dealt with optoelectronic devices designed to produce light for use in applications including displays, indicators, data storage and communications. The latter two of these applications also require the ability to detect the light once it

has been created. Much of the explosion of interest in semiconductor optoelectronic devices over the past 20 or so years has been particularly aimed at the communications market. This has put high demands on the ability to detect small optical signals being transmitted at very

fast data rates (> 40 Gbit/s). In this section, we consider the important aspects for detectors and describe some of the approaches that have been used.

37.6.1 Photodetector Requirements

There are many attributes important to the design of photodetectors. These can broadly be split into the following categories: sensitivity, speed, noise, physical size (footprint), reliability, temperature sensitivity, ease of use and finally cost. When deciding on a particular photodetector one must decide which of these factors is of most importance to the particular application of interest.

37.6.2 Photodetection Theory

The primary requirement of the band structure of semiconductor optical detectors is that the band gap of the active region is less than the photon energy to be absorbed. The electron-hole pair formed when the photon is absorbed is then separated by an applied or built-in field that exists in the active region. If, under the influence of the field, the electron and hole move apart by a distance, x , then the charge induced to flow in the external circuit connected to the detector is x/l . In a simple photoconductive detector consisting of a bulk semiconductor with conducting electrodes, l is the distance between the electrodes. In a p-n or p-i-n diode detector, l is the width of the depletion region.

Several factors need to be taken into account when considering the band structure of a detector. Firstly, the absorption depth of the light to be detected must be matched to the width of the active region, so that there is a high probability that the photogenerated carriers will be produced within the active region or less than the

minority-carrier diffusion length from it. Figure 37.30 shows the absorption coefficient as a function of energy for direct-gap GaAs and for indirect-gap Si. An absorption coefficient of 10^3 cm^{-1} , typical for Si, corresponds to an absorption depth of $10 \mu\text{m}$, while 10^4 cm^{-1} for GaAs corresponds to a $10\text{-}\mu\text{m}$ absorption depth. One must also consider the effect of the band structure on the dark current in the device, since this sets the lower limit on the strength of the signal that can be detected. There are three important effects that can give rise to the generation of electron-hole pairs even when no light is falling on the device. These are: a) thermal generation, b) band-to-band tunnelling and c) avalanche breakdown. Processes (a) and (b) are illustrated in Fig. 37.31. Thermal generation, which occurs even at zero electric field, can itself be conveniently divided into two parts. Firstly there is the part i_{diff} due to the diffusion of minority carriers into the field region. Secondly there is the current i_{gr} due to thermal generation within the field region. The thermal generation current increases strongly with temperature but is relatively independent of applied field, although i_{gr} increases slightly with reverse bias since this causes the depletion width to increase. Process (b), band-to-band tunnelling, is indicated by the horizontal arrow i_t in Fig. 37.31. At large reverse fields, electrons at the top of the valence band see a triangular-shaped potential barrier through which it is possible for them to tunnel into the conduction band. This quantum-mechanical tunnelling is relatively independent of temperature but is strongly dependent on the applied field. Process (c), avalanche breakdown, which occurs at very large electric fields, has to be avoided except in avalanche photodetectors, which will be described in some detail below. Avalanche breakdown involves carriers with initial ki-

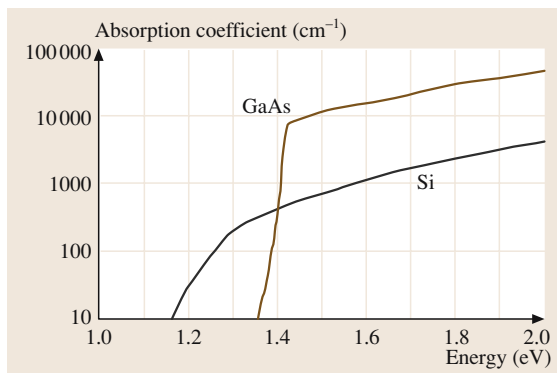


Fig. 37.30 Absorption coefficient as a function of energy for direct-gap GaAs and indirect-gap Si

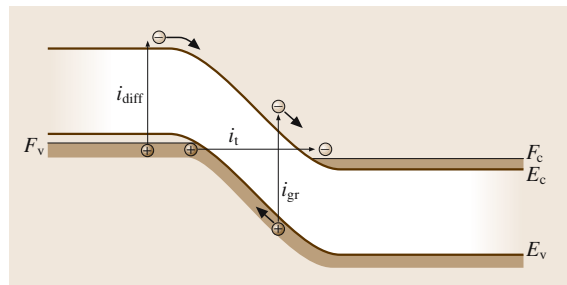


Fig. 37.31 Mechanisms contributing to the dark current in a reverse-biased p-i-n diode. Thermal generation has contributions from diffusion of minority carriers into the field region i_{diff} and from thermal generation within the field region i_{gr} . Zener (band-to-band) tunnelling is indicated by the horizontal arrow i_t

netic energies greater than the band-gap energy E_g away from the conduction-band minimum. This means that electrons in wide-band-gap III–V semiconductors will be at energies such that they are scattering between the central Γ minimum, the three X minima at the edges of the Brillouin zone in the $\langle 100 \rangle$ directions and the four L minima at the edges of the Brillouin zone in the $\langle 111 \rangle$ directions. Hydrostatic pressure measurements [37.53] showed that the effective ionisation energy $\langle E \rangle$ could be related to the average conduction-band energy simply by

$$\langle E \rangle = \frac{1}{8} (E_\Gamma + 3E_X + 4E_L) \quad (37.40)$$

where E_Γ , E_X and E_L are the energies of the Γ , X and L minima above the top of the valence band, respectively.

The first two effects giving rise to a dark current, described in (a) and (b) above, increase almost exponentially with decreasing band gap E_g . Therefore there exists an optimum value for the band gap E_g to be used to detect photons of a particular energy E_{ph} . Considering E_g to be the adjustable parameter, which can be determined by the alloy composition, as E_g is reduced below E_{ph} the absorption coefficient increases and light becomes more efficiently absorbed within the active region of the diode. However, once all the photons are absorbed, reducing E_g further adds nothing to the photocurrent but does continue to increase the dark current. Therefore, it is clear that the optimum band gap is just below that of the photon energy to be detected. The exact difference $E_{ph} - E_g$ depends upon the shape of the absorption edge of the material, the thickness of the active region and its position within the device.

37.6.3 Detectors with Internal Gain

In a photoconductive detector, gain can actually be achieved by adding impurities that trap one type of carrier, forming a centre with a low capture cross section for the other type of carrier. Assuming the trapped carrier is a hole, as the free electron reaches the anode another will be injected from the cathode and this process will continue until an electron recombines with the trapped hole. Under these circumstances x exceeds l and a gain of x/l is achieved: $x/l = \tau_r/t_t$, where t_t is the transit time of the electron between the electrodes and τ_r is the electron–hole recombination lifetime. The response time of such a system is $\tau_r = t_t x/l$ so that, as the gain increases, the speed of the device decreases. A related effect is obtained in a phototransistor. If, for example, we consider a photon being absorbed in the base or col-

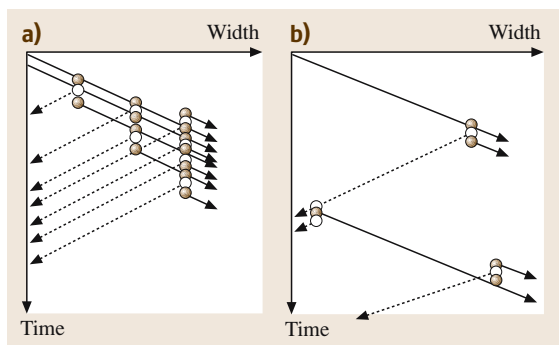


Fig. 37.32a,b Avalanche multiplication as a function of position and time where (a) the electron multiplication rate (solid line) is much greater than that for holes (dashed line). (b) when the electron and hole multiplication rates are equal, the current pulse continues for a long time

lector region of an n–p–n transistor then the presence of the hole, which will be effectively trapped in the base region, will cause the transit of electrons between the emitter and collector for the recombination lifetime of the hole. The gain of the system is then equal to the gain of the transistor, which is related to τ_r/t_t , where t_t is now the transit time of the electron from emitter to collector.

37.6.4 Avalanche Photodetectors

Interestingly, silicon as well as direct-band-gap materials can be used to produce efficient detectors, but for different reasons. In the direct-band-gap materials, photons with energies greater than the band gap are strongly absorbed by the creation of electron–hole pairs, leading to the generation in the detector of a primary photocurrent or photovoltage. In silicon, photons with energies just above the band gap are relatively weakly absorbed but the electrons generated can be greatly amplified by the process of avalanche multiplication. This occurs when a large reverse voltage is applied to the p–n or p–i–n junction of the detecting diode. Electrons drifting in the field can gain kinetic energy in excess of the band-gap energy E_g . It is then possible for the hot carrier to give up its energy and momentum in an impact ionisation process, resulting in the generation of an electron–hole pair. The generated carriers can, in turn, drift in the electric field and themselves generate further carriers. The band structure of silicon is particularly well suited for avalanche multiplication since impact ionisation can occur for hot electrons with conservation of energy and momentum when the electron’s kinetic energy is almost exactly equal to E_g [37.54]. By contrast

a hot hole in silicon requires an energy close to $3E_g/2$ before it can generate an electron–hole pair. Thus, in silicon, a large avalanche multiplication can be achieved with each electron that is optically generated and swept across the avalanche region. This situation is illustrated in Fig. 37.32a. Very large gains can be obtained but all of the carriers are swept out of the active region in a time equal or less than $t_e + t_h$, where t_e and t_h are the transit times taken by electrons and holes, respectively, to cross the active region.

In contrast, it is found that, in the majority of direct-band-gap materials, holes and electrons are able to produce impact ionisation at the same electric field. As a result, above this field, avalanching occurs for both types of carrier as they move in opposite directions and the carrier density escalates uncontrollably, leading to complete avalanche breakdown. This situation is illustrated in Fig. 37.32b. Gain, but also breakdown, occurs when the field is large enough to create an impact ionisation process just as the electron is about to leave the active region. If exactly the same happens for the hole, the situation can continue ad infinitum, leading to infinite gain but a current pulse that does not switch off. Of course, in fact, impact ionisation is a statistical process and the current pulse might stop after one electron transit or run away as described, leading to an unstable noisy detection system.

Silicon is therefore clearly a very useful material for the production of avalanche photodetectors (APDs). Unfortunately however, the detection bandwidth of silicon limits it from the visible spectrum up to $\approx 1\ \mu\text{m}$ in the near-infrared (IR). It is of significant technological importance to have fast high-gain low-noise detectors above this wavelength, in particular in the range $1.3\text{--}1.6\ \mu\text{m}$ for optical communications. The mid-infrared is also of importance due to the large number

of gas absorption bands in this range. Due to the effects discussed above, the performance of APDs in the near-mid-infrared region is poor when compared with silicon APDs. Recently, Adams proposed that APDs based upon InGaAsN may give rise to improved characteristics for detection in the near- and mid-IR [37.55]. As discussed earlier in this chapter, InGaAsN is already showing great promise for long-wavelength lasers due to the unusually large band-gap bowing brought about by small concentrations of nitrogen in GaAs. This clearly makes it a potential candidate for the manufacture of long-wavelength detectors. However, in addition to the large band-gap bowing, InGaAsN has an unusually large conduction-band effective mass derived from the dilute amount of nitrogen in the lattice. This has several effects of benefit to APDs. Firstly, the high effective mass inhibits electrons from gaining kinetic energy under an applied field. Secondly, the increased mass, and hence increased density of states, will increase the probability of electron scattering. As the electrons heat up and their energy approaches that of the N levels, they will be strongly scattered. Furthermore, if any electrons manage to reach the top of the E^- band, they have to be scattered across, or tunnel through, the energy gap to the next conduction (E^+) band before they can gain more energy. All of these effects prevent electrons from avalanching. However, since the valence band is largely unaffected by the presence of nitrogen, the holes may be accelerated by the field and avalanche. It has been proposed that such devices will have a multiplication of $\approx 10^4$ compared with < 10 in the best currently available APDs at these wavelengths. Furthermore, it is also predicted that the noise factor of these devices could be reduced by two orders of magnitude. Although yet to be demonstrated in practice, InGaAsN may in the future provide both high-quality detectors as well as emitters in the infrared.

37.7 Conclusions

This chapter introduced some of the most important optoelectronic devices in use today. It has discussed how these devices were developed based upon highly innovative work which has transformed semiconductor growth technology over the past 40 years. The requirements for lasing action in semiconductors were discussed as well as the ways in which

low-dimensional structures and strain may be used to improve the properties of devices such as lasers (both edge-emitting and surface-emitting), modulators, amplifiers and detectors. The chapter considered current state-of-the-art semiconductor laser devices and discussed emerging future directions in semiconductor optoelectronics.

References

- 37.1 N. Jnr. Holonyak, S. F. Bevacqua: Appl. Phys. Lett. **1**, 82 (1962)
- 37.2 A. Einstein: Phys. Z. **18**, 121 (1917)
- 37.3 M. G. A. Bernard, G. Duraffourg: Phys. Status Solidi **1**, 699 (1961)
- 37.4 M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill Jr., G. Lasher: Appl. Phys. Lett. **1**, 62 (1962)
- 37.5 R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, R. O. Carlson: Phys. Rev. Lett. **9**, 366 (1962)
- 37.6 T. M. Quist, R. H. Rediker, R. J. Keyes, W. E. Krag, B. Lax, A. L. McWhorter, H. J. Zeiger: Appl. Phys. Lett. **1**, 91 (1962)
- 37.7 H. Nelson, J. I. Pankove, F. Hawrylo, G. C. Dousmanis: Proc. IEEE **52**, 1360 (1964)
- 37.8 Zh. I. Alferov, V. M. Andreev, D. Z. Garbuzov, Yu. V. Zhilyaev, E. P. Morozov, E. L. Portnoi, V. G. Trofim: Sov. Phys. Semicond. **4**, 1573 (1971)
- 37.9 R. Dingle, C. H. Henry: US Patent 3982207 (1976)
- 37.10 A. R. Adams: Electron. Lett. **22**, 249 (1986)
- 37.11 E. Yablonovitch, E. O. Kane: J. Light. Technol. **LT-4**, 504 (1986)
- 37.12 Y. Arakawa, H. Sakaki: Appl. Phys. Lett. **40**, 939 (1982)
- 37.13 G. Liu, A. Stintz, H. Li, K. J. Malloy, L. F. Lester: Electron. Lett. **35**, 1163 (1999)
- 37.14 Y. Sakakibara, K. Furuya, K. Utaka, Y. Suematsu: Electron. Lett. **16**, 456 (1980)
- 37.15 M.-C. Amann, J. Buus: *Tunable Laser Diodes* (Artech House, Boston 1998)
- 37.16 K. Iga: IEEE J. Sel. Top. Quant. Electron. **6**, 1201 (2000)
- 37.17 J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, A. Y. Cho: Science **264**, 553 (1994)
- 37.18 F. Capasso, C. Gmachl, R. Paiella, A. Tredicucci, A. L. Hutchinson, D. L. Sivco, J. N. Baillargeon, A. Y. Cho, H. C. Liu: IEEE J. Sel. Top. Quant. Electron. **6**, 931 (2000)
- 37.19 T. Kojima, M. Tamura, H. Nakaya, S. Tanaka, S. Tamura, S. Arai: Jpn. J. Appl. Phys. **37**, 4792 (1998)
- 37.20 P. J. A. Thijs, L. F. Tiemeijer, J. J. Binsma, T. van Dongen: IEEE J. Quant. Electron. **30**, 477 (1994)
- 37.21 G. Jones, A. Ghiti, M. Silver, E. P. O'Reilly, A. R. Adams: IEE Proc. J. **140**, 85 (1993)
- 37.22 A. Valster, A. T. Meney, J. R. Downes, D. A. Faux, A. R. Adams, A. A. Brouwer, A. J. Corbijn: IEEE J. Sel. Top. Quant. Electron. **3**, 180 (1997)
- 37.23 E. P. O'Reilly: Semicond. Sci. Technol. **4**, 121 (1989)
- 37.24 T. Yamamoto, H. Nobuhara, K. Tanaka, T. Odagawa, M. Sugawara, T. Fujii, K. Wakao: IEEE J. Quant. Electron. **29**, 1560 (1993)
- 37.25 M. P. C. M. Krijn, G. W. 't Hooft, M. J. B. Boermans, P. J. A. Thijs, T. van Dongen, J. J. M. Binsma, L. F. Tiemeijer: Appl. Phys. Lett. **61**, 1772 (1992)
- 37.26 G. Jones, A. D. Smith, E. P. O'Reilly, M. Silver, A. T. R. Briggs, M. J. Fice, A. R. Adams, P. D. Greene, K. Scarrott, A. Vranic: IEEE J. Quant. Electron. **34**, 822 (1998)
- 37.27 A. Ghiti, M. Silver, E. P. O'Reilly: J. Appl. Phys. **71**, 4626 (1992)
- 37.28 S. J. Sweeney, A. F. Phillips, A. R. Adams, E. P. O'Reilly, P. J. A. Thijs: IEEE Phot. Tech. Lett. **10**, 1076 (1998)
- 37.29 Y. Zou, J. S. Osinski, P. Godzinski, P. D. Dapkus, W. Rideout, W. F. Sharfin, R. A. Logan: IEEE J. Quant. Elec. **29**, 1565 (1993)
- 37.30 G. P. Agrawal, N. K. Dutta: *Long-Wavelength Semiconductor Lasers* (Van Nostrand, New York 1986)
- 37.31 A. R. Adams, M. Asada, Y. Suematsu, S. Arai: Jap. J. Appl. Phys. **19**, L621 (1980)
- 37.32 R. Fehse, S. Jin, S. J. Sweeney, A. R. Adams, E. P. O'Reilly, H. Riechert, S. Illek, A. Yu. Egorov: Electron. Lett. **37**, 1518 (2001)
- 37.33 S. J. Sweeney, A. R. Adams, E. P. O'Reilly, M. Silver, P. J. A. Thijs: *Conference on Lasers and Electro-Optics* (IEEE, San Francisco 2000)
- 37.34 W. W. Lui, T. Yamanaka, Y. Yoshikuni, S. Seki, K. Yokoyama: Phys. Rev. B. **48**, 8814 (1993)
- 37.35 S. J. Sweeney, G. Knowles, T. E. Sale, A. R. Adams: Phys. Status Solidi B. **223**, 567 (2001)
- 37.36 J. I. Pankove: IEEE J. Quant. Electron. **QE-4**, 119 (1968)
- 37.37 E. P. O'Reilly, M. Silver: Appl. Phys. Lett. **63**, 3318 (1993)
- 37.38 N. Hatori, K. Otsubo, M. Ishida, T. Akiyama, Y. Nakata, H. Ebe, S. Okumura, T. Yamamoto, M. Sugawara, Y. Arakawa: *30th European Conference on Optical Communication* (IEEE, Stockholm 2004)
- 37.39 B. W. Hakki, T. I. Paoli: J. Appl. Phys. Lett. **46**, 1299 (1975)
- 37.40 A. I. Onischenko, T. E. Sale, E. P. O'Reilly, A. R. Adams, S. M. Pinches, J. E. F. Frost, J. Woodhead: IEE Proc. Optoelectron. **147**, 15 (2000)
- 37.41 T. E. Sale: *Vertical Cavity Surface Emitting Lasers* (Res. Stud., London 1995)
- 37.42 Y. Ohiso, C. Amano, Y. Itoh, H. Takenouchi, T. Kurokawa: IEEE J. Quant. Electron. **34**, 1904 (1998)
- 37.43 M. Kondow, S. Kitatani, S. Nakatsuka, M. C. Larson, K. Nakahara, Y. Yazawa, M. Okai, K. Uomi: IEEE J. Sel. Top. Quant. Electr. **3**, 719 (1997)
- 37.44 T. Kitatani, K. Nakahara, M. Kondow, K. Uomi, T. Tanaka: Jpn. J. Appl. Phys. **39**, L86 (2000)
- 37.45 K. D. Choquette, J. F. Klem, A. J. Fischer, O. Blum, A. A. Allerman, I. J. Fritz, S. R. Kurtz, W. G. Breiland, R. Sieg, K. M. Geib, J. W. Scott, R. L. Naone: Electron. Lett. **36**, 1388 (2000)
- 37.46 G. Steinle, H. Riechert, A. Yu. Egorov: Electr. Lett. **37**, 93 (2001)

- 37.47 R. Averbeck, G. Jaschke, L. Geelhaar, H. Riechert: *19th International Semiconductor Laser Conference 2004, Conference Digest* (IEEE, Piscataway, NJ 2004)
- 37.48 M. A. Wistey, S. R. Bank, H. B. Yuen, L. L. Goddard, J. S. Harris: *J. Vac. Sci. Tech. B* **22**, 1562 (2004)
- 37.49 L. F. Tiemeijer, P. J. A. Thijs, T. van Dongen, R. W. M. Slootweg, J. M. M. van der Heijden, J. J. M. Binsma, M. P. C. M. Krijn: *Appl. Phys. Lett.* **62**, 826 (1993)
- 37.50 F. Devaux, S. Chelles, A. Ougazzaden, A. Mircea, J. C. Harmand: *Semicond. Sci. Technol.* **10**, 887 (1995)
- 37.51 M. Silver, P. D. Greene, A. R. Adams: *Appl. Phys. Lett.* **67**, 2904 (1995)
- 37.52 I. K. Czajkowski, M. A. Gibbon, G. H. B. Thompson, P. D. Greene, A. D. Smith, M. Silver: *Electron. Lett.* **30**, 900 (1994)
- 37.53 J. Allam, A. R. Adams, M. A. Pate, J. S. Roberts: *Inst. Phys. Conf. Ser.* **112**, 375 (1990)
- 37.54 A. R. Adams, J. Allam, I. K. Czajkowski, A. Ghiti, E. P. O'Reilly, W. S. Ring: Strained-layer lasers and avalanche photodetectors. In: *Condensed Systems of Low Dimensionality*, ed. by J. L. Beeby (Plenum, New York 1991) p. 623
- 37.55 A. R. Adams: *Electr. Lett.* **40**, 1086 (2004)

Liquid Crystal

38. Liquid Crystals

This chapter outlines the basic physics, chemical nature and properties of liquid crystals. These materials are important in the electronics industry as the electro-optic component of flat-panel liquid-crystal displays, which increasingly dominate the information display market.

Liquid crystals are intermediate states of matter which flow like liquids, but have anisotropic properties like solid crystals. The formation of a liquid-crystal phase and its properties are determined by the shape of the constituent molecules and the interactions between them. While many types of liquid-crystal phase have been identified, this Chapter focuses on those liquid crystals which are important for modern displays.

The electro-optical response of a liquid crystal display depends on the alignment of a liquid-crystal film, its material properties and the cell configuration. Fundamentals of the physics of liquid crystals are explained and a number of different displays are described.

In the context of materials, the relationship between the physical properties of liquid crystals and their chemical composition is of vital importance. Materials for displays are mixtures of many liquid-crystal compounds carefully tailored to optimise the operational behaviour of the display. Our current understanding of how chemical structure determines the physical properties is outlined, and data for typical liquid-crystal compounds are tabulated. Some key

38.1	Introduction to Liquid Crystals	917
38.1.1	Calamitic Liquid Crystals	919
38.1.2	Chiral Liquid Crystals	921
38.1.3	Discotic Liquid Crystals	923
38.2	The Basic Physics of Liquid Crystals	924
38.2.1	Orientational Order	924
38.2.2	Director Alignment	925
38.2.3	Elasticity	926
38.2.4	Flexoelectricity	928
38.2.5	Viscosity	929
38.3	Liquid-Crystal Devices	931
38.3.1	A Model Liquid-Crystal Display: Electrically Controlled Birefringence Mode (ECB)	932
38.3.2	High-Volume Commercial Displays: The Twisted Nematic (TN) and Super-Twisted Nematic (STN) Displays	935
38.3.3	Complex LC Displays and Other Cell Configurations	937
38.4	Materials for Displays	940
38.4.1	Chemical Structure and Liquid-Crystal Phase Behaviour ..	942
38.4.2	The Formulation of Liquid-Crystal Display Mixtures	942
38.4.3	Relationships Between Physical Properties and Chemical Structures of Mesogens	943
	References	949

references are given, but reference is also made to more extensive reviews where additional data are available.

38.1 Introduction to Liquid Crystals

Liquid crystals have been known for almost 120 years but it is only in the last 30 years or so that their unique application in display devices has been recognised. Now they are seen as extremely important materials having made possible the development of thin screens for use with personal computers (PCs) and in televisions. In

fact a wide range of different liquid-crystal (LC) display devices has been developed. The common feature for each of these is that the optical characteristics of the display are changed on application of an electric field across a thin liquid-crystal film. The process causing this change is associated with a variation in the

macroscopic organisation of the liquid crystal within the cell. The liquid crystal is, therefore, strictly behaving as a molecular material and not an electronic one. Nonetheless the display itself is closely integrated with electronic components. Since liquid crystals may be unfamiliar to those concerned with conventional electronic materials, this section begins with an introduction to liquid crystals and the compounds that form them. The following section describes the basic physics for liquid crystals which are needed to understand their use in display devices. The functioning of the most important displays is described in Sect. 38.3, which makes contact with the basic physics outlined in Sect. 38.2. The liquid crystal materials used in display devices are discussed in the final section, where the necessary optimisation of a wide range of properties is addressed.

The majority of chemical compounds can exist in three states of matter, namely crystal, liquid or gas, each with its defining characteristics. There is a fourth state known as a liquid crystal and, as the name suggests, this state has characteristics of both crystals and liquids. Thus a liquid crystal flows when subject to a stress, like a liquid, but certain of its properties are anisotropic, like a crystal. This macroscopic behaviour, often used to identify the phase, implies that at the microscopic or molecular level the material has an element of long-range orientational order together with some translational disorder at long range. It is this combination of order and disorder that makes liquid

crystals so fascinating and gives them their potential for applications, especially in the field of electro-optic displays.

A variety of different classes of materials are known to form liquid crystals at some point on their phase diagram [38.1]. These include organic materials where the liquid crystal is formed, on heating, between the crystal and isotropic liquid phases. Such materials are known as thermotropic liquid crystals and are the subject of this Chapter. Another class is formed by amphiphilic organic materials in which part of the constituent molecules favours one solvent, normally water, while the other part does not. When the amphiphile is dissolved in the water, the molecules form aggregates which then interact to give the liquid-crystal phase, the formation of which is largely controlled by the concentration of amphiphile. These are known as lyotropic liquid crystals; they underpin much of the surfactant industry, although they are not used in displays and so will not be considered further. Colloidal dispersions of inorganic materials such as clays can also form liquid-crystal phases depending on the concentration of the colloidal particles. Solutions of certain organic polymers also exhibit liquid-crystal phases and, like the colloidal systems, the solvent acts to increase the separation between polymer chains but does not significantly affect their state of aggregation. A prime example of such a system is the structural polymer Kevlar, which for the same weight is stronger than steel; it is formed by processing a nematic solution of the polymer.

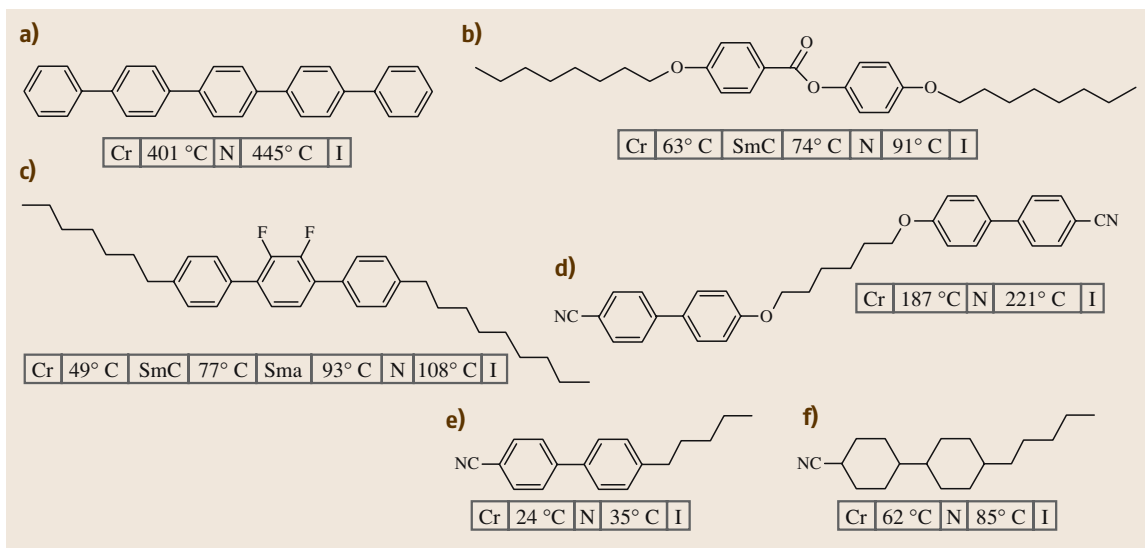


Fig. 38.1 The molecular structures for a selection of compounds which form calamitic liquid crystals

The following sections return to thermotropic liquid crystals and describe the molecular organisation within the phases, mention some of their properties and briefly indicate the relationship between the phase and molecular structures.

38.1.1 Calamitic Liquid Crystals

In view of the anisotropic properties of liquid crystals, it seems reasonable that a key requirement for their formation is that the molecules are also anisotropic. This is certainly the case, with the majority of liquid crystals having rod-like molecules, such as those shown in Fig. 38.1. One of the simplest nematogenic rod-like molecules is *p*-quinquephenyl (Fig. 38.1a), which is essentially rigid. However, flexible alkyl chains can also be attached at one or both ends of the molecule (Fig. 38.1b and c) or indeed in the centre of the molecule (Fig. 38.1d), and rigid polar groups such as a cyano (Fig. 38.1d–f) may be attached at the end of the molecule. The rigid part is usually constructed from planar phenyl rings (Fig. 38.1a–e) but they can be replaced by alicyclic rings (Fig. 38.1f) which enhance the liquid crystallinity of the compound. The term calamitic, meaning rod-like, is applied to the phases that they form. There are, in fact, many different calamitic liquid-crystal

phases but we shall only describe those which are of particular relevance to display applications.

At an organisational level the simplest liquid crystal is called the nematic and in this phase the rod-like molecules are orientationally ordered, but there is no long-range translational order. A picture showing this molecular organisation, obtained from a computer simulation of a Gay–Berne mesogen [38.2] is given in Fig. 38.2b. The molecular shape is ellipsoidal and the symmetry axes of the ellipsoids tend to be parallel to each other and to a particular direction known as the director.

In contrast there is no ordering of the molecular centres of mass, except at short range. The essential difference between the nematic and isotropic phases (Fig. 38.2a) is the orientational order, which is only short range in the isotropic liquid. At a macroscopic level the nematic phase is characterised by its high fluidity and by anisotropy in properties such as the refractive index. The anisotropic properties for a nematic have cylindrical symmetry about the director, which provides an operational definition of this unique axis and is the optic axis for the phase. The anisotropy in the refractive index combined with the random director distribution results in the turbidity of the phase, which contrasts with the transparency of the isotropic liquid. This on its own would not be sufficient to identify the liquid crystal as a nematic phase but identification is possible from the optical texture observed under a polarising microscope. These textures act as fingerprints for the different liquid-crystal phases. An example of such a texture for a nematic phase is shown in Fig. 38.3a; it is created by the anisotropy or birefringence in the refractive index combined with a characteristic distribution of the director in the sample.

The next level of order within liquid crystal phases is found for the smectic A phase. Now, in addition to the long-range orientational order, there is translational order in one dimension, giving the layered structure shown in Fig. 38.2c [38.2]. The director associated with the orientational order is normal to the layers. Within a layer there is only short-range translational order as in a conventional liquid. In this structure the layer spacing is seen to be comparable but slightly less than the molecular length, as found experimentally for many smectic phases.

At a macroscopic level the layer structure means that the fluidity of a smectic A phase is considerably less than for a nematic phase. The properties are anisotropic and the birefringence is responsible for the turbidity of the phase, as found for a nematic. However, under a polar-

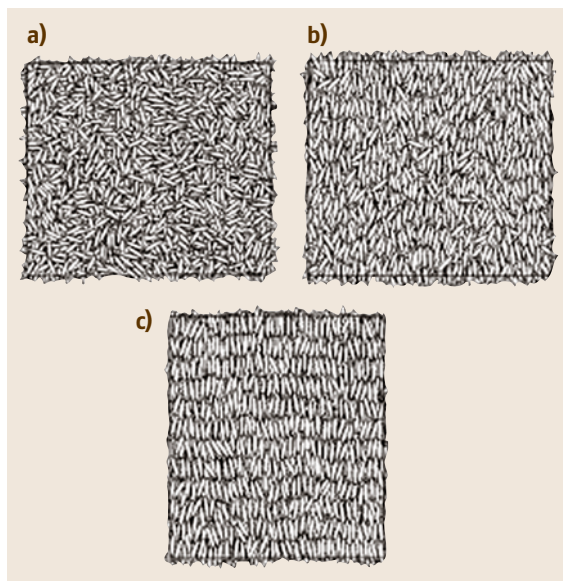


Fig. 38.2a–c The molecular organisation in (a) the isotropic phase (b) the nematic phase and (c) the smectic A phase obtained from the simulation of a Gay–Berne calamitic mesogen



Fig. 38.3a–c Typical optical textures observed with a polarising microscope for (a) nematic, (b) smectic A and (c) columnar liquid-crystal phases

ising microscope the optical texture is quite different to that of a nematic, as is apparent from the focal conic fan texture shown in Fig. 38.3b.

A variant on the smectic A is the smectic C phase. The essential difference to the smectic A phase is that the director in that smectic C phase is tilted with respect to the layer normal. The defining characteristic of the smectic C phase is then the tilt angle, which is taken as the angle between the director and the layer normal. This tilt in the structure reduces the symmetry of the phase to the point group C_{2h} in contrast to $D_{\infty h}$ for nematic and smectic A phases. This lowering in symmetry naturally influences the symmetry of the properties. The fluidity of the smectic C phase is comparable to that of a smectic A phase. However, the optical texture can be quite different and it has elements similar to a nematic phase and to a smectic A; the focal conic fan structure is less well defined and is said to be broken. The nematic-like features result from the fact that the tilt direction of the director is not correlated between the smectic layers and so it adopts a distribution analogous to that of the director in a nematic phase.

The molecular factors which influence the ability of a compound to form a liquid-crystal phase have been well studied both experimentally [38.3] and theoretically [38.4]. Consider the simple nematic as formed, for example, by *p*-quinquephenyl (Fig. 38.1a); this melts at 401 °C to form the nematic phase, which then undergoes a transition to the isotropic phase at 445 °C; the transition temperatures are denoted by T_{CrN} and T_{NI} , respectively. The very high value of T_{NI} , which is a measure of the stability of the nematic phase, is attributed to the large length-to-breadth ratio of *p*-quinquephenyl. In contrast *p*-quaterphenyl, formed by the removal of just one of the five phenyl rings, does not exhibit a liquid-crystal phase at atmospheric pressure, even though its shape anisotropy is still relatively large. This occurs be-

cause, on cooling, the isotropic liquid freezes before the transition to the nematic phase can occur. Indeed many compounds with anisotropic molecules might be expected to form liquid-crystal phases, but do not because of their high melting points. As a consequence the molecular design of liquid crystals needs to focus not only on increasing the temperature at which the liquid crystal–isotropic transition occurs but also on lowering the melting point. One way by which this can be achieved is to attach flexible alkyl chains to the end of the rigid core (Fig. 38.1). In the crystal phase the chain adopts a single conformation but in a liquid phase there is considerable conformational disorder, and it is the release of conformational entropy on melting that lowers the melting point. The addition of the chain also affects the nematic–isotropic transition temperature, which alternates as the number of atoms in the chain passes from odd to even. This odd–even effect is especially dramatic when the flexible chain links two mesogenic groups (Fig. 38.1d) to give what is known as a liquid-crystal dimer [38.5]. The odd–even effect is particularly marked because the molecular shapes for the dimers with odd and even spacers differ significantly on average, being bent and linear, respectively.

The attachment of alkyl chains to the rigid core of a mesogenic molecule has another important consequence, as it tends to promote the formation of smectic phases (Fig. 38.1b and c). The reason that the chains lead to the formation of such layered structures is that, both energetically and entropically, the flexible chains prefer not to mix with the rigid core, and so by forming a layer structure they are able to keep apart. Indeed it is known that biphenyl (a rigid rod-like structure) is not very soluble in octane (a flexible chain). The lack of compatibility of the core and the chain increases with the chain length and so along a homologous series it is those members with long alkyl

chains that form smectic phases. For example, 4-pentyl-4'-cyanobiphenyl (Fig. 38.1e) forms only a nematic phase whereas the longer-chain homologue, 4-decyl-4'-cyanobiphenyl only exhibits a smectic A phase. To obtain a tilted smectic phase such as a smectic C there clearly needs to be a molecular interaction which favours an arrangement for a pair of parallel molecules that is tilted with respect to the intermolecular vector. Such a tilted structure can be stabilised by electrostatic interactions; for example by off-axis electric dipoles (Fig. 38.1b and c) [38.6] or with a quadrupolar charge distribution.

The ability of a compound to form a liquid crystal is not restricted to just one phase. The delicate balance of the intermolecular interactions responsible for the various liquid-crystal phases means that transitions between them can result from modest variations in temperature. This is apparent for the 4,4'-dialkyl-2',3'-difluoroterphenyl shown in Fig. 38.1c, which on cooling the isotropic liquid forms nematic, smectic A and smectic C phases; such a compound is said to be polymorphic. Materials which form even more liquid-crystal phases are known [38.1]. The occurrence of several liquid-crystal phases in a single system can be of value in processing the material for display applications.

38.1.2 Chiral Liquid Crystals

The mesogenic molecules that have been considered so far are achiral in the sense that the molecule is superimposable on its mirror image. Molecules may also be

chiral in that they are not superimposable on their mirror images; this chirality can result from the tetrahedral arrangement of four different groups around a single carbon atom. This is illustrated in Fig. 38.4a which shows such an arrangement together with its mirror image; these are known as enantiomers. The presence of a chiral centre will certainly change the nature of the interactions between the molecules and it is relevant to see whether the chiral interactions might not influence the structure of the liquid-crystal phases exhibited by the material. From a formal point of view it might be expected that the molecular chirality of a mesogen should be expressed through the symmetry of the liquid-crystal structure. This proves to be the case, provided no other interactions oppose the chiral deformation of the liquid-crystal phase. In fact the first liquid crystal to be discovered [38.7] was chiral; this was cholesteryl benzoate where the cholesteryl moiety contains many chiral centres. The structure of the liquid crystal is nematic-like in that there is no long-range translational order but there is long-range orientational order. However, the difference between this phase and a nematic formed from achiral materials is that the director is twisted into a helix. The helix may twist in a left-handed or a right-handed sense and these structures, shown in Fig. 38.4b, are mirror images of each other. The phase structure is certainly chiral and so is known as a chiral nematic, although originally it was called a cholesteric phase. The symbol for the phase is N^* , where the asterisk indicates that the phase has a chiral structure.

The helical structure is characterised by the pitch of the helix p which is the distance along the helix axis needed for the director to rotate by 2π . Since the directions parallel and antiparallel to the director are equivalent the periodicity of the chiral nematic is $p/2$. For many chiral nematics the helical pitch is comparable to the wavelength of visible light. This, together with the periodic structure of the phase, means that Bragg reflection from a chiral nematic will be in the visible region of the spectrum and so this phase will appear coloured with the wavelength of the reflected light being related to the pitch of the helix. This pitch is sensitive to temperature, especially when the chiral nematic phase is followed by a smectic A. This sensitivity has been exploited in the thermochromic application of chiral nematics, where the reflected colour of the phase changes with temperature [38.1].

The chirality of cholesteryl benzoate clearly results from the chiral centres present in the mesogenic molecule. However, the chirality can also be introduced

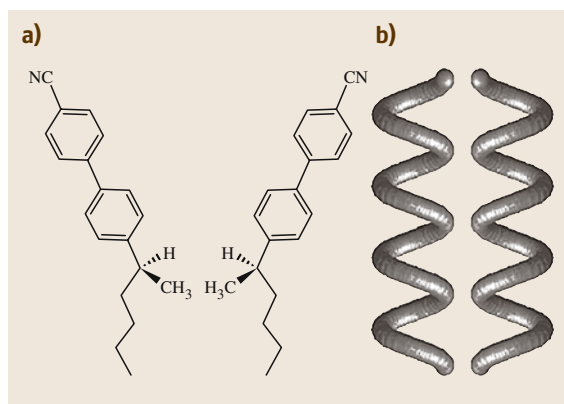


Fig. 38.4 (a) A chiral mesogenic molecule, (R) 2-[4'-cyano-4-biphenyl]-hexane, together with its mirror image, (S) 2-[4'-cyano-4-biphenyl]-hexane. (b) The left- and right-handed helical organisation of the director for a chiral nematic

indirectly to a mesogen by simply adding a chiral dopant, which does not need to be mesogenic. The mixture will be chiral and this is sufficient to lead to a chiral nematic. The pitch of this mixture depends on the amount of the dopant and the inverse pitch, p^{-1} , proves to be proportional to its concentration. The handedness of the helix induced by the dopant will depend on its stereochemical conformation and will be opposite for the two enantiomers. Accordingly, if both enantiomers are present in equal amounts, i. e. as a racemic mixture, then doping a nematic with this will not convert it to a chiral nematic.

Chiral smectic phases are also known in which the director adopts a helical structure as a result of introducing molecular chirality into the material either as a dopant or as an intrinsic part of the mesogenic molecule. The chiral smectic C phase, denoted SmC^* , provides an appropriate example with which to illustrate the structure of such phases. In an achiral smectic C phase the tilt direction of the director changes randomly from layer to layer, analogous to the random director orientation in an achiral nematic. For the chiral smectic C phase, as might be anticipated, the tilt direction of the director rotates in a given sense, left-handed or right-handed, and by a small, fixed amount from layer to layer. Other structural features of the smectic C phase remain unchanged. Thus, the director of the chiral smectic C phase has a he-

lical structure with the helix axis parallel to the layer normal. The pitch of the helix is somewhat smaller than that of the associated chiral nematic phase. The magnitude of the pitch is inversely related to the tilt angle of the smectic C, and since this grows with decreasing temperature so the pitch decreases. The reduced symmetry, C_2 , of the SmC^* phase leads to the introduction of a macroscopic electrical polarisation [38.8]. This is of potential importance for the creation of fast-switching displays.

The ability of the SmC^* phase to adopt a helical structure results from the fact that the tilt direction for the director acts in an analogous manner to the director in a nematic, and importantly that the layer spacing is preserved in the helical structure. In marked contrast there are strong forces inhibiting the creation of a twisted structure for a smectic A composed of chiral molecules. The director is normal to the layers and so the creation of a twisted structure would require a variation in the layer thickness but this has a high energy penalty associated with it. Accordingly many of the smectic A phases formed from chiral molecules have the same structure as those composed of achiral molecules. There are, however, exceptions and these occur when the chiral interactions are especially strong and, presumably, the translational order of the layers is small. Under such

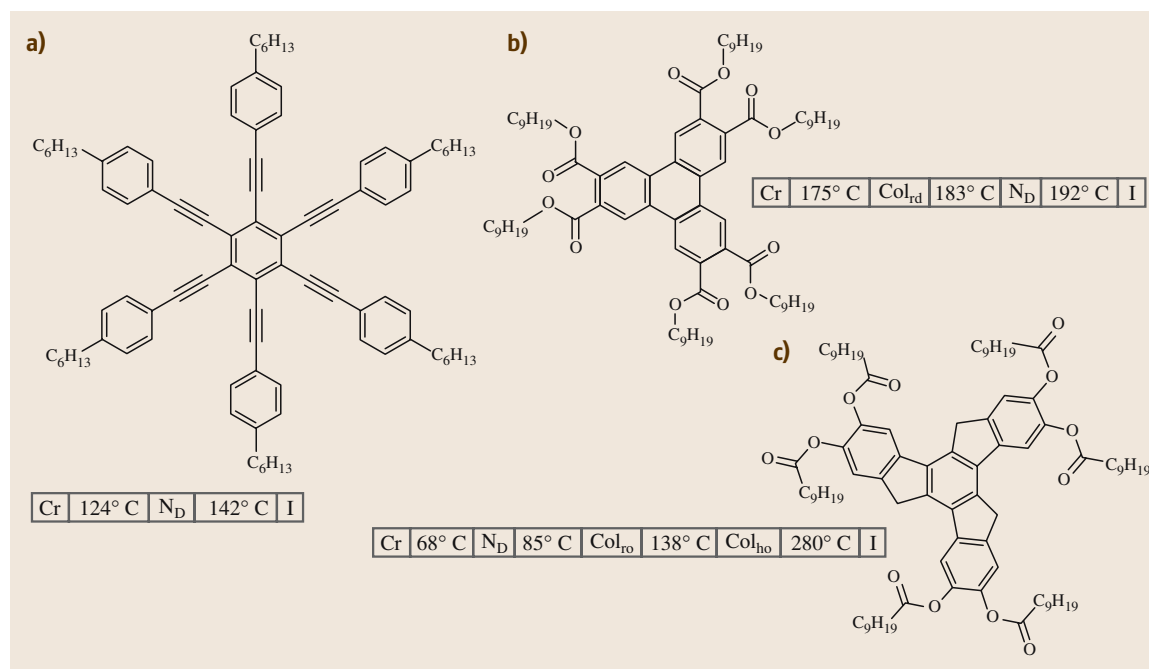


Fig. 38.5 A selection of molecular structures for compounds that form discotic liquid crystals

conditions the smectic A structure is partially destroyed, creating small SmA-like blocks about 1000 Å wide, separated by screw dislocations [38.9, 10]. These defects in the organisation allow the directions for the blocks to rotate coherently to give a chiral helical structure; the pitch of the helix is found to be larger than that in an analogous chiral nematic phase. This chiral phase is just one example of a liquid-crystal structure stabilised by defects; it is known as a twist grain-boundary phase and denoted by TGBA*. The letter A indicates that the director is normal to the layers in the small smectic blocks; there is a comparable phase in which the director is tilted, denoted by TGBC*.

38.1.3 Discotic Liquid Crystals

The key requirement for the formation of a liquid crystal is an anisotropic molecule, as exemplified by the rod-like molecules described in the previous sections. However, there is no reason why disc-like molecules should not also exhibit liquid-crystal phases. Nonetheless, it was not until 1977 that the first example of a thermotropic

liquid crystal formed from disc-like molecules was reported [38.11]. Since that time several liquid-crystal phases have been identified and these phases are known collectively as discotic liquid crystals. The range of compounds that exhibits these phases is now extensive and continues to grow [38.12], although it does not match the number that form calamitic liquid crystals. The molecular structures of three compounds which form discotic liquid crystals are shown in Fig. 38.5.

As for rod-like molecules, the simplest liquid-crystal phase formed by disc-like molecules is the nematic, usually denoted N_D , where the D indicates the disc-like nature of the molecules. Within the nematic phase, shown in Fig. 38.6a, the molecular centres of mass are randomly distributed and the molecular symmetry axes are orientationally correlated. The nematic structure is the same as for that formed from rod-like molecules, the only difference being that the symmetry axes which are orientationally ordered are the short axes for the discs and the long axes for the rods. The point symmetry of the discotic and calamitic nematic is the same, namely $D_{\infty h}$. The discotic nematic is recognised in the same way as the calamitic nematic; that is it flows like a normal fluid and its anisotropy is revealed by a characteristic optical texture analogous to that shown in Fig. 38.3a. In fact, the refractive index of a discotic nematic along the director is smaller than that perpendicular to it, which is the opposite to that for a calamitic nematic. It has been suggested that this difference may be of value in display devices [38.13] but this concept has not as yet been commercialised.

The other class of discotic liquid crystals possesses some element of long-range translational order and these are known as columnar phases, two examples of which are sketched in Fig. 38.6b and c. The disc-like molecules are stacked face-to-face into columns. A single column has a one-dimensional structure, and as such is not expected to exhibit long-range translational order, although this can result from interactions between neighbouring columns in the liquid-crystal phase. The column–column interactions will result in the columns being aligned parallel to each other; these interactions will also determine how the columns are packed. When the discs are orthogonal to the column axis the cross section is essentially circular and so the columns pack hexagonally, as shown in Fig. 38.6b. The symbol given to this phase is Col_{hd} , where h denotes hexagonal packing of the columns and d indicates that the arrangement along the column is disordered. The point group symmetry of this phase is D_{6h} . The disc-like molecules may also be tilted with respect to the column axis, giving

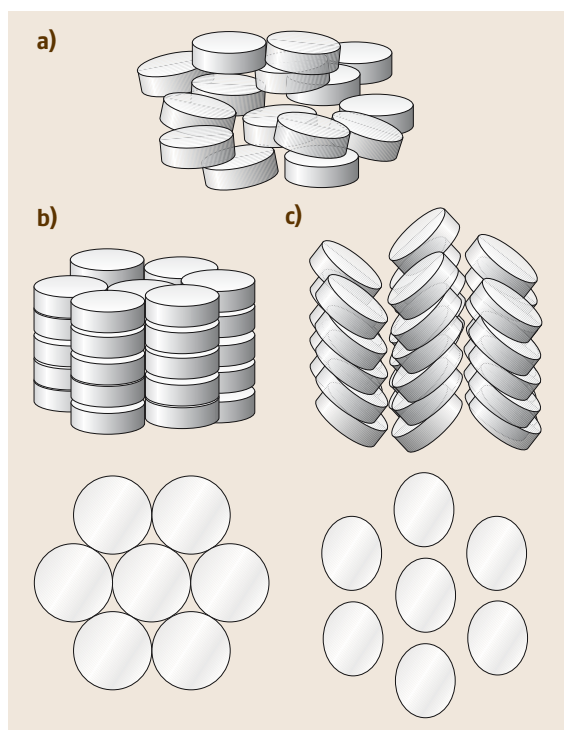


Fig. 38.6a–c The molecular organisation in discotic liquid-crystal phases, (a) nematic, (b) hexagonal columnar and (c) rectangular columnar

an elliptical cross section to the columns. As a result the columns are packed on a rectangular lattice; there are four possible arrangements and just one of these is indicated in Fig. 38.6c. In general the mnemonic used to indicate a rectangular columnar phase is Col_{rd}. The point group symmetry of the rectangular columnar phase is D_{2h} and the extent to which the structure deviates from that of the Col_{hd} phase will depend on the magnitude of the tilt angle within the column. The columnar phases can be identified from their optical textures and an example of one is shown in Fig. 38.3c.

The columnar phases have potential electronic applications because of the inhomogeneity of the molecules that constitute them; i.e. the central part is aromatic while the outer part is aliphatic. As a result of the overlap between the π -orbitals on the centres of neighbouring discs it should be possible for electrical conduction to take place along the core of the column. This should occur without leakage into adjacent columns because of the insulation provided by the alkyl chains. It should also be possible to anneal these molecular wires because of their liquid-crystal properties [38.14]. This and the ability to avoid defects in the columns which can prevent electronic conduction in crystals mean that the columnar phase has many potential advantages over non-mesogenic materials. In addition, discotic systems are also used as compensating films to improve the optical characteristics for some liquid-crystal displays.

At a molecular level the factors that are responsible for the formation of the discotic liquid-crystal phases are similar to those for calamitic systems. Thus, the molecular design should aim to increase the liquid

crystal–isotropic transition temperature while decreasing the melting point. The latter is certainly achieved by attaching flexible alkyl chains to the perimeter of the rigid disc. The creation of the columnar phases should be relatively straightforward provided the central core is both planar and large. Then, because of the strong attractive forces between the many atoms in the rigid core the molecules will wish to stack face-to-face in a column. The formation of the columns will also be facilitated by the flexible alkyl chains attached to the core. Clearly then it may prove to be difficult to create the nematic phase before the columnar phase is formed unless the disc–disc interactions can be weakened. One way in which this can be achieved is by destroying the planarity of the core, for example, by using phenyl rings attached to the molecular centre so that they can rotate out of the plane (Fig. 38.5a). It is to be expected that the columnar phases should occur below the nematic phase, corresponding to an increase in order with decreasing temperature. This is usually observed, for example, for the hexasubstituted triphenylenes (Fig. 38.5b). However, the truxene derivatives, with long alkyl chains on the perimeter, (Fig. 38.5c) exhibit quite unusual behaviour. For these compounds the crystal melts to form a discotic nematic and then at a higher temperature a columnar phase appears. This deviates from the expected sequence, and because the nematic phase appears at a lower temperature than the columnar phase it is usually referred to as a re-entrant nematic. The occurrence of a re-entrant phase is often attributed to a conformational change which strengthens the molecular interactions with increasing temperature thus making the more ordered phase appear at higher temperatures.

38.2 The Basic Physics of Liquid Crystals

38.2.1 Orientational Order

The defining characteristic of a liquid crystal is the long-range order of its constituent molecules. That is, for rod-like molecules, the molecular long axes tend to align parallel to each other even when separated by large distances. The molecules tend to be aligned parallel to a particular direction known as the director and denoted by \mathbf{n} . This is an apolar vector, that is $\mathbf{n} = -\mathbf{n}$, because the nematic does not possess long-range ferroelectric order. The properties of the nematic phase are cylindrically symmetric about the director, which provides a macroscopic definition of this. The anisotropy of the

properties results from the orientational order and the extent of this is commonly defined [38.15] by

$$S = \left\langle (3 \cos^2 \beta - 1) / 2 \right\rangle, \quad (38.1)$$

although other definitions are possible [38.16]. Here β is the angle made by a molecule with the director and the angular brackets indicate the ensemble average. In the limit of perfect order S is unity while in the isotropic phase S vanishes. The temperature dependence of S is shown in Fig. 38.7 for the nematogen, 4,4'-dimethoxyazobenzene; this behaviour is typical of most nematic liquid crystals. At low temperatures S is about 0.6 and then decreases with increasing

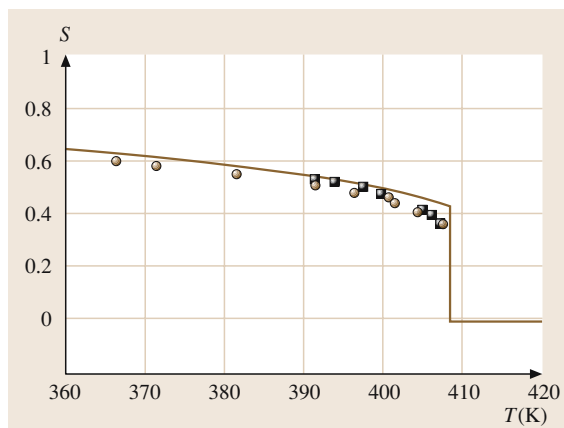


Fig. 38.7 The temperature variation of the orientational order parameter, S , for 4,4'-dimethoxyazobenzene; the different symbols indicate results determined with different techniques

temperature, reaching about 0.3 before it vanishes discontinuously at the nematic–isotropic transition, in keeping with the first-order nature of this transition. It is also found, both experimentally and theoretically, that the orientational order of different nematic liquid crystals is approximately the same provided they are compared at corresponding temperatures, either the reduced, T/T_{NI} , or shifted, $T_{\text{NI}} - T$, temperatures [38.16]. Since many properties of liquid crystals are related to the long-range orientational order these also vary with temperature especially in the vicinity of the transition to the isotropic phase.

38.2.2 Director Alignment

The director in a bulk liquid crystal is distributed randomly unless some constraint is applied to the system; a variety of constraints can be employed and two of these are of special significance for display applications. One of them is an electric field and because of the inherent anisotropy in the dielectric permittivity of the liquid crystal the director will be aligned. The electric energy density controlling the alignment is given by [38.17]

$$U_{\text{elec}} = -\varepsilon_0 \Delta\varepsilon (\mathbf{n} \cdot \mathbf{E})^2 / 2. \quad (38.2)$$

Here ε_0 is the permittivity of a vacuum and the scalar product $\mathbf{n} \cdot \mathbf{E}$ is $E \cos \theta$, where E is the magnitude of the field \mathbf{E} and θ is the angle between the director and the field, $\Delta\varepsilon$ is the anisotropy in the dielectric tensor

$$\Delta\varepsilon = \varepsilon_{||} - \varepsilon_{\perp}, \quad (38.3)$$

where the subscripts denote the values parallel ($||$) and perpendicular (\perp) to the director. If the dielectric anisotropy is positive then the director will be aligned parallel to the electric field and, conversely, if $\Delta\varepsilon$ is negative, then the director is aligned orthogonal to the field. The molecular factors that control the sign of $\Delta\varepsilon$ will be discussed in Sect. 38.4. Of course, intense electric fields are also able to align the molecules in an isotropic phase but what is remarkable about a nematic liquid crystal is the very low value of the field needed to achieve complete alignment of the director. Thus for a bulk nematic free of other constraints the electric field necessary to align the director is typically about 30 kV/m, although the value does clearly depend on the magnitude of the dielectric anisotropy. This relatively small value results because of the long-range orientational correlations which mean that the field acts, in effect, on the entire ensemble of molecules and not just single molecules.

The other constraint, essential for display devices, is the interaction between the director and the surface of the container [38.18]. At the surface there are two extreme arrangements for the director. One is with the director orthogonal to the surface, the so-called homeotropic alignment. In the other the director is parallel to a particular direction in the surface; this is known as the uniform planar alignment. The type of alignment depends on the way in which the surface has been treated. For example, for a glass surface coated with silanol groups a polar liquid crystal will be aligned homeotropically, while to achieve this alignment for a non-polar nematic the surface should be covered with long alkyl chains. In these examples the direct interaction of a mesogenic molecule with the surface produces an orthogonal alignment which is then propagated by the long-range order into the bulk. To achieve uniform planar alignment of the director the surface is coated with a polymer, such as a polyimide, which on its own would result in planar alignment. To force the director to be parallel to a particular direction in the surface the polymer is rubbed which aligns the director parallel to the direction of rubbing. There is still some uncertainty about the mechanism responsible for uniform planar alignment. It might result from alignment of the polymer combined with anisotropic intermolecular attractions with the mesogenic molecules, although it had been thought [38.19] to have its origins in surface grooves and the elastic interactions which are described later.

The energy of interaction between the surface and the director clearly depends on the nature of the surface treatment and the particular nematic. *Rapini* and

Popular [38.20] have suggested the following simple form for the surface energy density

$$U_S = -A(\mathbf{n} \cdot \mathbf{e})^2/2, \quad (38.4)$$

where A is the anchoring energy and \mathbf{e} is the easy axis or direction along which the director is aligned. Clearly it has an analogous form to that for the anisotropic interaction between the nematic and an electric field (38.2) but is essentially phenomenological. The anchoring energy is determined to be in the range 10^{-7} – 10^{-5} J/m² [38.21]. The upper value corresponds to strong anchoring in that typical values of the electric field would not change the director orientation at the surface. In contrast the lower value is associated with weak anchoring and here the director orientation at the surface can be changed by the field.

38.2.3 Elasticity

As the name suggests, a liquid crystal has some properties typical of crystals and others of liquids. The elastic properties of crystals should, therefore, be reflected in the behaviour of liquid crystals. Here it is the director orientation which is the analogue of the atomic positions in a crystal. In the ground state of a nematic liquid crystal the director is uniformly aligned. However, the elastic torques, responsible for this uniform ground state, are weak and, at temperatures within the nematic range, the thermal energy is sufficient to perturb the director configuration in the bulk. This perturbed state can take various forms depending on a combination of fac-

tors but, whatever the form, it can be represented as a sum of just three fundamental distortion modes. These are illustrated in Fig. 38.8 and are the splay, twist and bend deformations [38.22]. At a more formal level these modes are also shown in terms of the small deviations of the director from its aligned state at the origin as the location from this is varied. For example, for the twist deformation away from the origin along the x -axis there is a change in the y -component of the director, n_y , and the displacement along the y -axis causes a change in n_x . The magnitude of the twist deformation is measured by the gradients $\partial n_y/\partial x$ and $\partial n_x/\partial y$.

The energy needed to stabilise a given distortion of the director field is clearly related to the extent of the deformation via the gradients. This distortion energy density for a bulk nematic liquid crystal is given by continuum theory [38.22] as

$$f = \left[K_1(\nabla \cdot \mathbf{n})^2 + K_2(\mathbf{n} \cdot \nabla \times \mathbf{n})^2 + K_3(\mathbf{n} \times \nabla \times \mathbf{n})^2 \right]/2, \quad (38.5)$$

where the terms $\nabla \cdot \mathbf{n}$, $\mathbf{n} \cdot \nabla \times \mathbf{n}$ and $\mathbf{n} \times \nabla \times \mathbf{n}$ correspond to the splay, twist and bend deformations, respectively. The contribution each makes to the free energy is determined by the proportionality constants K_1 , K_2 and K_3 , which are usually known as the Frank elastic constants for splay, twist and bend, respectively. They are small, typically 5×10^{-12} N, and their small magnitude explains why the thermal energy is able to distort the uniform director arrangement so readily. The elastic constants are not in fact constant but vary with temperature

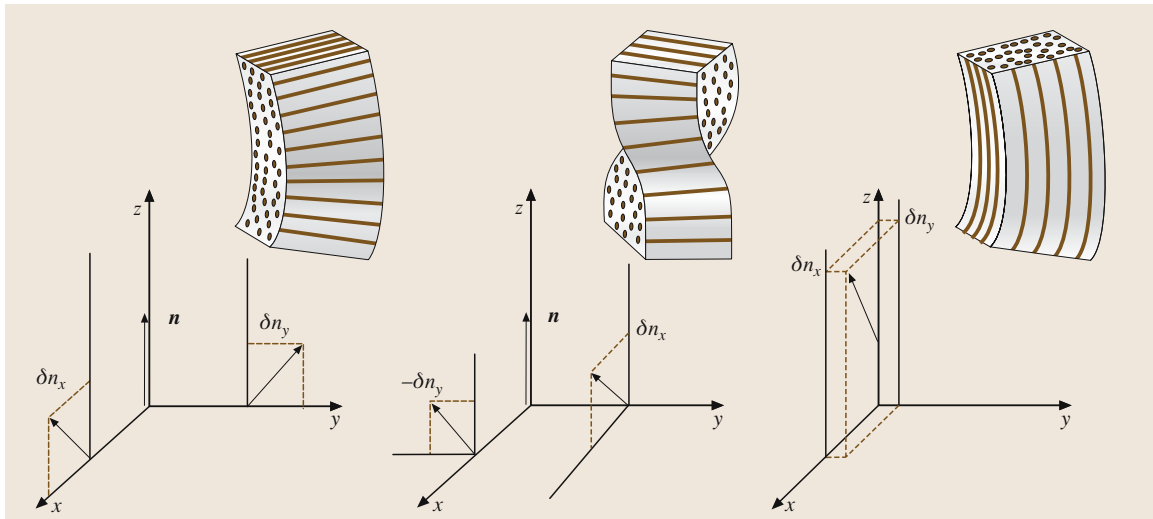


Fig. 38.8a–c The three fundamental deformations for the nematic director (a) splay, (b) twist and (c) bend

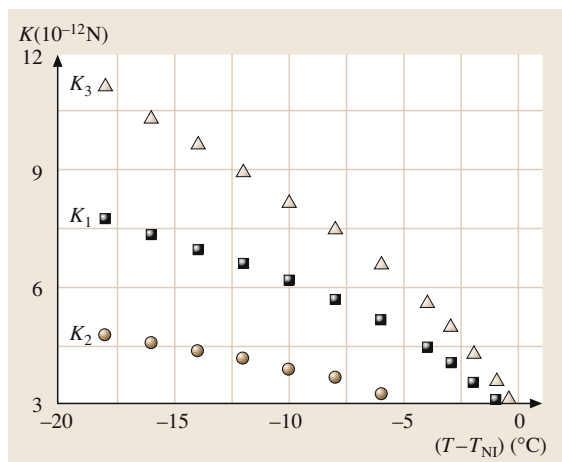


Fig. 38.9 The temperature dependence of the three elastic constants, K_1 , K_2 and K_3 , for the nematic 5CB

and nor are they equal as the results for 4-pentyl-4'-cyanobiphenyl (5CB) shown in Fig. 38.9 demonstrate. The twist elastic constant is seen to be the smallest while the largest is the bend elastic constant. This means that it is easiest to induce a twist deformation in a nematic while a bend deformation is the most difficult to create. All three elastic constants decrease with increasing temperature in keeping with the decreasing order as the transition to the isotropic phase is approached. Like the orientational order the elastic constants vanish discontinuously at the first-order nematic–isotropic transition.

The continuum theory is especially valuable in predicting the behaviour of display devices and this is illustrated by considering one of the ingenious experiments devised by *Fréedericksz* to determine the elastic constants [38.23]. In these a thin slab of nematic is confined between two glass plates with a particular director configuration, either uniform planar or homeotropic produced by surface forces. These forces control the director orientation just at the two surfaces and the alignment across the slab is propagated by the elastic interactions. A field is then applied which will move the director away from its original orientation and the variation of the director orientation with the field strength provides the elastic constant. To employ the continuum theory in order to describe the experiment it is necessary to add the field energy density (38.2) to the elastic free-energy density in (38.5). The director configuration is then obtained by integrating the free-energy density over the volume of the sample, and minimising this, subject to the surface constraints. The geometry of the experiment is shown in Fig. 38.10 and provided the dielectric anisotropy, $\Delta\epsilon$,

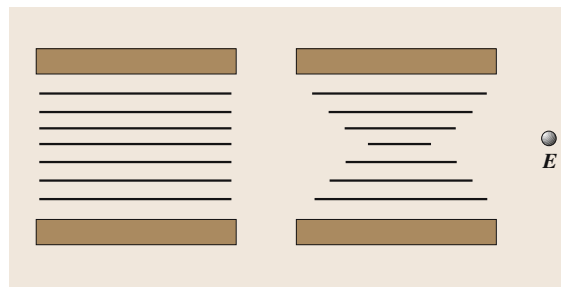


Fig. 38.10a,b The geometry for the Fréedericksz experiment used to determine the twist elastic constant, K_2 ; (a) in zero field and (b) above the threshold value

is positive the director will move from being orthogonal to the field to being parallel to it. However, the extent of this twist deformation will vary across the cell, being greatest at the centre and zero at the surfaces, in the limit of strong anchoring. The dependence of the director orientation, at the centre of the slab, with respect to the electric field is shown in Fig. 38.11. As the field strength is increased from zero the director orientation remains unchanged until a threshold value, E_{th} , is reached when the angle between the director and the field starts to decrease continuously. At very high values of the field the director at the centre of the slab tends to be parallel to the field. This behaviour can be understood in the following simple terms. Below the threshold field the elastic energy exceeds the electrical energy and so the director retains its uniform planar alignment. Above the threshold field strength the elastic energy is less than

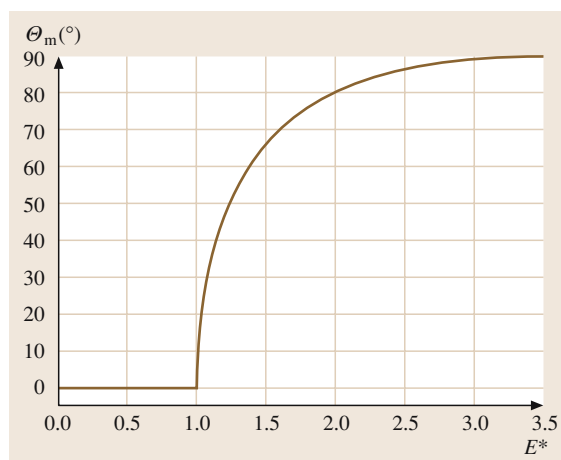


Fig. 38.11 The dependence of the director orientation θ_m at the centre of a nematic slab on the scaled field strength, $E^* = E/E_{th}$

the electrical energy and so the director begins to move to be parallel to the electric field. The threshold electric field is predicted to be

$$E_{\text{th}} = (\pi/l)\sqrt{K_2/\epsilon_0\Delta\epsilon}, \quad (38.6)$$

where l is the slab thickness. Once the threshold field has been measured and provided the dielectric anisotropy is known, the twist elastic constant would be available. Other Fréedericksz experiments with analogous expressions for the threshold field lead to the determination of the splay and bend elastic constants.

For real display applications (Sect. 38.3) the director alignment at the surface deviates from either uniform planar or homeotropic alignment. This deviation is known as a surface pre-tilt and is illustrated in Fig. 38.6a for near-uniform planar alignment with the tilt direction on the two surfaces differing by 180° . The cell with this arrangement has what is known as antiparallel alignment, so named because of the difference in the tilt direction caused by the direction of rubbing on the surfaces being antiparallel. In zero field, therefore, the director is uniformly aligned across the cell but tilted with respect to the x -axis set in the surface. A continuum theory calculation analogous to the case when the pre-tilt angle, θ_0 , is zero allows the dependence of the director orientation θ_m in the centre of the cell to be determined as a function of the strength of the field applied across the cell. The results of these calculations,

for a nematic with positive $\Delta\epsilon$, are shown in Fig. 38.12b as a function of the scaled field strength, E/E_{th} , where E_{th} is the threshold field for zero pre-tilt. The theoretical dependence for a tilt angle of zero is analogous to that considered for the twist deformation. In other words, below the threshold field the director is parallel to the easy axis and then above this threshold the director moves to become increasingly parallel to the field. When there is a surface pre-tilt the behaviour is quite different and this is especially apparent for a pre-tilt angle of 10° (Fig. 38.12b). In zero field the angle θ_m made by the director with the x -axis is 10° and, as the field increases, so does θ_m . The rate of increase grows as the threshold field is approached and then is reduced as E/E_{th} increases beyond unity. Comparison of this behaviour with the conventional Fréedericksz experiment (Fig. 38.12b) shows that a pre-tilt of 10° has a significant effect on the way in which the director orientation changes with the field strength. Indeed, even for a pre-tilt of just 2° , there is a pronounced difference in behaviour in the vicinity of the threshold field. This clearly has important implications for the accurate determination of the elastic constants [38.20]. It also shows how unique the behaviour is when the surface pre-tilt angle is zero.

Chiral nematics are often employed in liquid-crystal display devices. Locally, their structure is analogous to a nematic but the director is twisted into a helical structure. The continuum theory for the chiral nematic must, therefore, be consistent with the helical ground state structure of the phase. To achieve this, a constant is added to the twist term in the elastic free-energy density. Thus (38.5) for a nematic becomes

$$f = \left[K_1(\nabla \cdot \mathbf{n})^2 + K_2(\mathbf{n} \cdot \nabla \times \mathbf{n} - 2\pi/p)^2 + K_3(\mathbf{n} \times \nabla \times \mathbf{n})^2 \right] / 2 \quad (38.7)$$

for a chiral nematic, where p is the pitch of the helix.

38.2.4 Flexoelectricity

Another property of solids which is mimicked by liquid crystals is piezoelectricity. For solids this is the generation of a macroscopic electrical polarisation as a result of the deformation of certain ionic materials. It is to be expected, therefore, that deformation of the director distribution for a liquid crystal will create a macroscopic polarisation; this proves to be the case [38.24] and the phenomenon is known as flexoelectricity. The origin of flexoelectricity can be understood in the following way. It is generally assumed that mesogenic molecules are

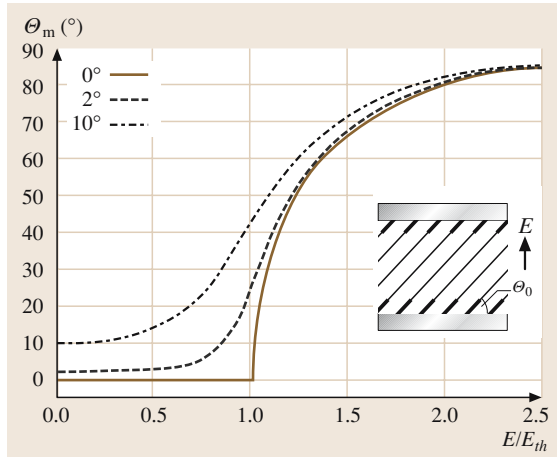


Fig. 38.12 (a) The director alignment in a cell with pre-tilt at the two surfaces assembled so that the rubbing directions are antiparallel. (b) The electric field dependence of the director orientation θ_m at the centre of the cell for values of the pre-tilt angle of 0° (solid line), 2° (dashed line) and 10° (dash-dotted line)

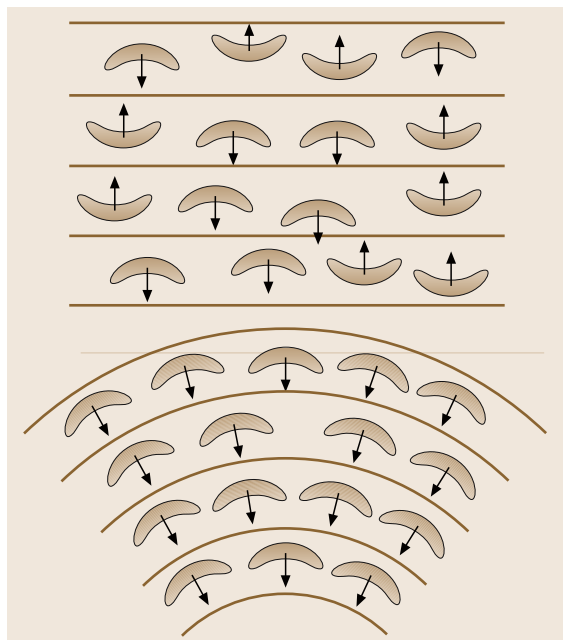


Fig. 38.13a,b The orientational distribution of bent molecules in (a) a uniformly aligned nematic and (b) one subject to a bend deformation; the x -axes are denoted by arrows

cylindrically symmetric but examination of the molecular structure of real mesogens shows that this is not the case. The molecules are asymmetric and at one extreme can be thought of as bent, as shown in Fig. 38.13; here z is the molecular long axis and x is the axis bisecting the bond angle. When the director is uniformly aligned the z -axis will be parallel to the director and the x -axis will be randomly arranged orthogonal to it (Fig. 38.13). If the director is now subject to a bend deformation the molecular long axis will still tend to be parallel to the local director, however, the x -axis will tend to align parallel to a direction in the plane formed by the bent director. This change in the distribution function for the x -axis will introduce polar order into the system and if there is an electrical dipole moment along the x -axis then the deformed nematic will exhibit a macroscopic polarisation, \mathbf{P} . A similar argument shows that, if the molecule is wedge-shaped, a splay deformation of the director distribution will also induce a polarisation in the nematic [38.24].

The magnitude of \mathbf{P} , the induced dipole per unit volume, clearly depends on the extent of the deformation in the director distribution. In the linear response regime the induced polarisation for mesogenic molecules of

arbitrary asymmetry is given by [38.25]

$$\mathbf{P} = e_1 \mathbf{n} \nabla \cdot \mathbf{n} + e_3 \mathbf{n} \times \nabla \times \mathbf{n}, \quad (38.8)$$

where the vectors $\mathbf{n} \nabla \cdot \mathbf{n}$ and $\mathbf{n} \times \nabla \times \mathbf{n}$ represent the splay and bend deformations, respectively. Since the polarisation per unit volume is also a vector then the proportionality constants, e_1 and e_3 , are scalars. These are known as the splay and bend flexoelectric coefficients, respectively, and their dimensions are Cm^{-1} . The determination of individual flexoelectric coefficients is challenging [38.26], however, it seems that $(e_1 + e_3)$ is of the order of 10^{-12} – 10^{-11} Cm^{-1} , although from the previous discussion their magnitude should vary amongst the mesogens because of their dependence on the molecular asymmetry and the size and location of the dipole moment. The molecular model proposed to understand the flexoelectricity of nematics suggests that, for rod-like molecules, devoid of asymmetry, the flexoelectric coefficients should vanish, but that does not seem to be the case. Indeed, it has been proposed that polarisation can result even for rod-like molecules if they possess an electrostatic quadrupole moment [38.27]. This is not inconsistent with the polarisation predicted by (38.8) which follows from the reduced symmetry of a nematic with splay and bend deformations of the director [38.24]. Although the existence of flexoelectricity is of considerable fundamental interest the inverse effect in which the director is deformed from a uniform state by the application of an electric field is of relevance for liquid-crystal displays (Sect. 38.3). This deformation occurs because the polarisation induced by the director deformation can couple with the applied electric field and so stabilise the deformation [38.26]. Since the coupling is linear in the electric field then reversal of the field will simply reverse the deformation; a novel bistable device based solely on this reversal is described in Sect. 38.3.

38.2.5 Viscosity

For a nematogen the fluidity of the nematic phase is comparable to that of the isotropic phase appearing at a higher temperature. It is this fluidity, similar to that of a liquid, which is responsible, in part, for the display applications of nematics. An indication of the fluidity of conventional liquids is provided by a single viscosity coefficient, η , measured from the flow of the liquid subject to an applied stress in a viscometer. The flow behaviour of a nematic is made more complex by its defining long-range orientational order and the resultant anisotropy of the phase.

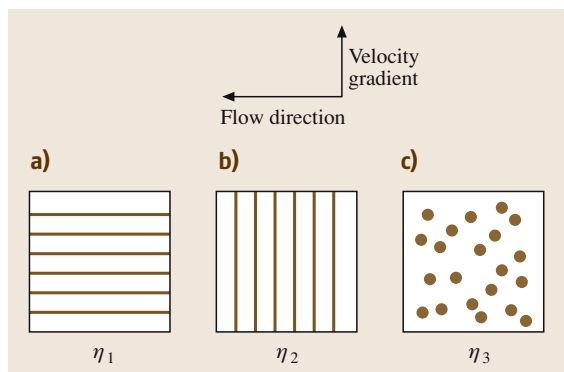


Fig. 38.14a–c The principal flow geometries of the Miesowicz experiments with the director pinned (a) parallel to the flow direction, (b) parallel to the velocity gradient and (c) orthogonal to both the flow direction and velocity gradient

This complexity can be appreciated at a practical level in terms of the Miesowicz experiments to determine the viscosity coefficients of a nematic [38.28]. In these experiments it is helpful to consider flow in a viscometer with a square cross section such that there is a velocity gradient orthogonal to the direction of flow. For a nematic, flow through the viscometer will now depend on the orientation of the director with respect to these two axes. A magnetic field is employed to align the director along a particular axis and it must be sufficiently strong that flow does not perturb the director alignment. There are three relatively simple flow geometries and these are shown in Fig. 38.14: (a) with the director parallel to the flow direction, (b) with the director parallel to the velocity gradient, and (c) with the director orthogonal to both the flow direction and velocity gradient. The viscosity coefficients were denoted by *Miesowicz* as (a) η_1 , (b) η_2 , and (c) η_3 , although other notation has been proposed in which η_1 and η_2 are interchanged [38.29]. The three viscosity coefficients are clearly expected to differ given the anisotropy of the nematic phase and these differences are found to be large. They are illustrated for the room-temperature nematogen 4-methoxybenzylidene-4'-butylaniline (MBBA) [38.30] in Fig. 38.15 where the viscosities are plotted against temperature.

It is immediately apparent that flow is easiest when the director is parallel to the direction of flow, as might have been anticipated. Conversely flow of the nematic is most difficult when the director is orthogonal to the flow direction but parallel to the velocity gradient. The intermediate viscosity, η_3 , occurs when the director is orthogonal to both the flow direction and the velocity

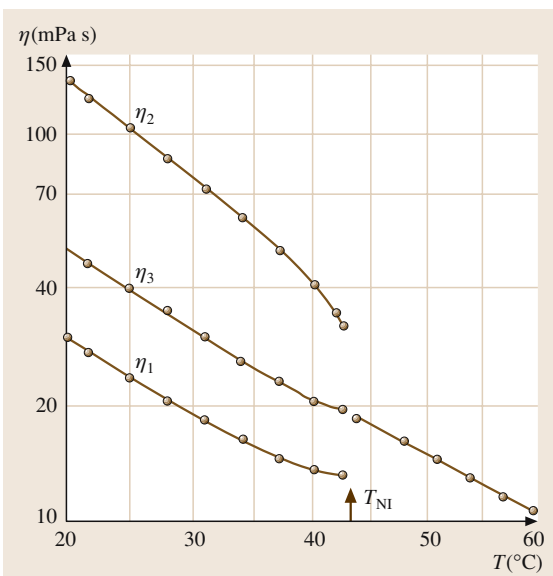


Fig. 38.15 The temperature dependence of the three Miesowicz viscosity coefficients, η_1 , η_2 , and η_3 , for the nematic phase of MBBA; the viscosity coefficient for the isotropic phase is also shown

gradient. Here it is seen that this third viscosity coefficient for the nematic is very similar to the viscosity coefficient for the isotropic phase extrapolated to lower temperatures.

The viscosity coefficient $\eta(\theta, \phi)$ for an arbitrary orientation (θ, ϕ) of the director with respect to the flow direction and velocity gradient can be related to the director orientation. It might be expected that this orientation dependence would involve just the three Miesowicz viscosity coefficients, rather like the transformation of a second-rank tensor from its principal axis system. In fact this is not quite the case and a fourth viscosity coefficient η_{12} needs to be introduced. The orientation dependence is then [38.28]

$$\eta(\theta, \phi) = \eta_1 \cos^2 \theta + \left(\eta_2 + \eta_{12} \cos^2 \theta \right) \times \sin^2 \theta \cos^2 \phi + \eta_3 \sin^2 \theta \sin^2 \phi; \quad (38.9)$$

we see that for the principal orientations $(0, \pi/2)$, $(\pi/2, 0)$ and $(\pi/2, \pi/2)$ the expression gives η_1 , η_2 , and η_3 , as required. The optimum director orientation with which to determine η_{12} is with the director at 45° to both the flow direction and the velocity gradient; the viscosity coefficient for this is

$$\eta(\pi/4, 0) = \eta_{12}/4 + (\eta_1 + \eta_2)/2. \quad (38.10)$$

The value of η_{12} determined for MBBA proves to be significantly smaller than the other three viscosity coefficients.

The four viscosity coefficients have been defined at a practical level in terms of flow in which the director orientation is held fixed. The converse of these experiments, in which the director orientation is changed in the absence of flow, allows the definition of the fifth and final viscosity coefficient. This is known as the rotational viscosity coefficient; it is denoted by the symbol γ_1 and plays a major role in determining the response times of display devices (Sect. 38.4). To appreciate the significance of γ_1 , it is helpful, as for the other four viscosity coefficients, to consider an experiment with which to measure it. In this an electric field is suddenly applied an angle θ to a uniformly aligned director, then providing the dielectric anisotropy is positive the director orientation will be changed and rotates towards the field direction. The electric torque responsible for the alignment is given by

$$\Gamma_{\text{elec}} = -\varepsilon_0 \Delta \varepsilon \sin 2\theta / 2, \quad (38.11)$$

which is the derivative of the electric energy in (38.2). The rotation of the director is opposed by the viscous torque

$$\Gamma_{\text{visc}} = \gamma_1 d\theta / dt, \quad (38.12)$$

which is linear in the rate at which the director orientation changes, with the proportionality constant being γ_1 . Provided the only constraint on the director is the electric field and provided, $\theta_0 \leq 45^\circ$, the director will move as a monodomain so that the elastic terms vanish. The inertial term for a nematic is small and so the

movement of the director is governed by the equation in which the two torques are balanced. That is, the electric torque causing rotation is balanced by the viscous torque opposing it; this gives

$$\gamma_1 d\theta / dt = -\varepsilon_0 \Delta \varepsilon E^2 \sin^2 \theta / 2. \quad (38.13)$$

The solution to this differential equation is

$$\tan \theta = \tan \theta_0 \exp(-t/\tau), \quad (38.14)$$

where θ_0 is the initial orientation of the director with respect to the electric field and τ is the relaxation time

$$\tau = \gamma_1 / \varepsilon_0 \Delta \varepsilon E^2. \quad (38.15)$$

Measurement of the time-dependent director orientation allows τ to be determined and from this γ_1 , given values of $\Delta \varepsilon$. For MBBA the rotational viscosity coefficient is found to be slightly less than η_2 and to parallel its temperature dependence [38.31].

The five independent viscosity coefficients necessary to describe the viscous behaviour of a nematic have been introduced in a pragmatic manner by appealing to experiments employed to measure these coefficients. However, the viscosity coefficients can be introduced in a more formal way as has been shown by Ericksen and then by Leslie in their development of the theory for nematodynamics [38.32]. The Leslie–Ericksen theory in its original form contained six viscosity coefficients, but subsequently Parodi has shown, using the Onsager relations, that there is a further equation linking the viscosity coefficients, thus reducing the number of independent coefficients to five [38.33]. These five coefficients are linearly related to those introduced by reference to specific experiments.

38.3 Liquid-Crystal Devices

The idea to use liquid crystals as electro-optic devices goes back to the early days of liquid-crystal research. In 1918 Björnstål, a Swedish physicist, demonstrated that the intensity of light transmitted by a liquid crystal could be varied by application of an electric field [38.34]. As optical devices of various types became established in the first decades of the 20th century, mainly in the entertainment industry, ways of controlling light intensity became important to the developing technologies. One device that soon found commercial application was the Kerr cell shutter, in which an electric field caused the contained fluid (usually nitrobenzene) to become birefringent. Placing such a cell between crossed polarisers

enabled a beam of light to be switched on and off very rapidly. The first report of liquid crystals being of interest for electro-optic devices was in 1936, when the Marconi Company filed a patent [38.35] which exploited the high birefringence of nematic liquid crystals in an electro-optic shutter. However, it was another 35 years before commercial devices became available which used the electro-optic properties of liquid crystals. The long interruption to the development of liquid-crystal devices can be attributed to the lack of suitable materials. We shall see in this Section how the physical properties of liquid crystals determine the performance of devices.

The optical properties of liquid crystals are exploited in displays, although the operational characteristics of such devices also depend crucially on many other physical properties (Sect. 38.2). Since the devices to be described all depend on the application of an electric field, their operation will be influenced by electrical properties such as dielectric permittivity and electrical conductivity. There is a range of electro-optic effects that can be used in devices, and the precise manner in which the properties of the liquid-crystal materials affect the device behaviour depends on the effect and the configuration of the cell. Thus there is not a single set of ideal properties that can define the best liquid-crystal material, rather the material properties have to be optimised for a particular application. In this section, different devices will be described, and their dependence on different properties will be outlined. The way in which the material properties can be adjusted for any application will be discussed in the final section of this Chapter.

38.3.1 A Model Liquid-Crystal Display: Electrically Controlled Birefringence Mode (ECB)

If an electric field is applied across a film of planar-aligned liquid crystal having a positive dielectric anisotropy, then the director of the liquid crystal will tend to align along the electric field. Thus the director will rotate into the field direction, and the optical retardation or birefringence of the film will change.

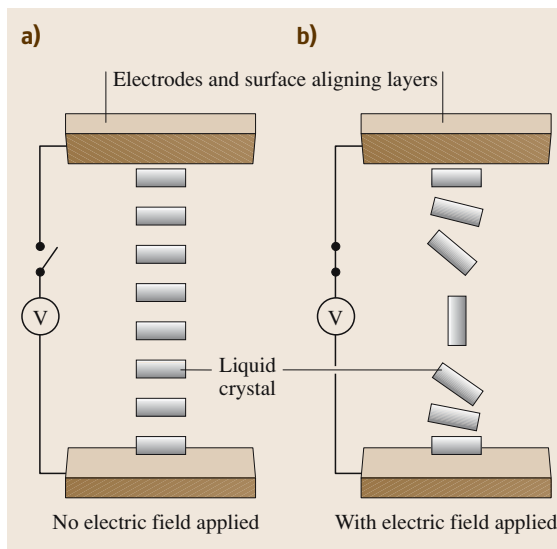


Fig. 38.16 Electrically controlled birefringence (ECB) cell

If the film is placed between crossed polarisers, then the change in optical retardation will be observed as a change in the intensity of transmitted light. The configuration for an ECB-mode display is schematically illustrated in Fig. 38.16. This illustrates the principal components of a liquid-crystal display. The liquid crystal is contained between two glass plates that have been coated with a transparent conducting layer, usually an indium–tin oxide alloy. Such treatment allows the application of an electric field across the liquid-crystal film, which can then be viewed along the field direction. For most applications, the surfaces of the electrodes are treated so that a particular director orientation is defined at the surface. This can be achieved in a variety of ways, depending on the desired surface director orientation. For the ECB-mode display under consideration, the director alignment should be parallel to the glass substrates and along a defined direction, which is the same on both surfaces. The standard technique to produce this alignment is to coat the glass plates with a thin (0.5 nm) layer of polyimide, which is mechanically rubbed in a particular direction. The cell is then placed between crossed polarisers, the extinction directions of which make angles of $\pm 45^\circ$ with the surface-defined director orientation.

The unperturbed state of the liquid-crystal film will be determined by the surfaces which contain it. If these have been treated in such a way that the liquid-crystal director is parallel to the surface along a particular direction, then the film will act as an optical retardation plate. Thus incident polarised light will, in general, emerge elliptically polarised, and there will be a phase retardation between components of the light wave parallel to the fast and slow axes of the retardation plate. For electromagnetic waves polarised at $\pm 45^\circ$ to the director the phase retardation ϕ , in radians, will be determined by the intrinsic birefringence of the liquid crystal ($\Delta n = n_e - n_o$), the film thickness ℓ and the wavelength of the light λ :

$$\phi = \frac{2\pi\ell}{\lambda}(n_e - n_o); \quad (38.16)$$

n_e and n_o are respectively the extraordinary (slow) and ordinary (fast) refractive indices of the liquid crystal (assuming that Δn is positive). If the emergent elliptically polarised light passes through a second polariser, crossed with respect to the incident polarisation direction, only a proportion of the incident intensity will be transmitted. The normalised intensity of light transmitted by a pair of crossed polarisers having a birefringent element between them, the axis of which is at $\pm 45^\circ$ to

the extinction directions of the polarisers is given by

$$T = \left(\frac{1}{2}\right) \sin^2 \frac{\phi}{2} \quad (38.17)$$

(the factor of one half appears for incident unpolarised light – if the light is polarised, as from a laser, then the factor is one).

Thus the initial appearance of the cell will be brightest if the cell thickness, birefringence and wavelength are chosen to give ϕ equal to π , 3π , 5π etc. It is normal to select the cell thickness to give $\phi = \pi$, and under these conditions the display is known as *normally white*. It is possible, though less satisfactory, to configure the display so that it operates in a *normally black* state, corresponding to a phase retardation of a multiple of 2π .

Application of an electric field causes the director orientation to change such that the optical retardation of the cell decreases to zero, and hence the cell becomes non-transmitting, at least in the normally white configuration. Under these circumstances the optical retardation across the cell becomes

$$\phi = \frac{2\pi}{\lambda} \int_0^{\ell} \{n_e[\theta(z)] - n_o\} dz, \quad (38.18)$$

where the effective extraordinary refractive index $n_e[\theta(z)]$ depends on the angle $[90^\circ - \theta(z)]$ between the director and the field and is a function of position z in the cell. This effective index is given by

$$\frac{1}{n_e^2[\theta(z)]} = \frac{\sin^2[\theta(z)]}{n_o^2} + \frac{\cos^2[\theta(z)]}{n_e^2}; \quad (38.19)$$

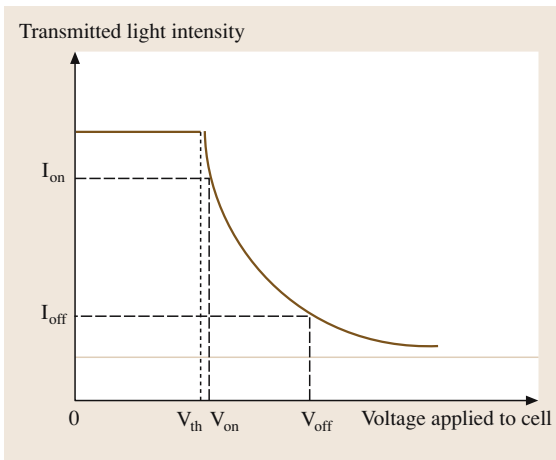


Fig. 38.17 Variation of optical transmission with voltage for an ECB cell between crossed polarisers

when the director is along the field direction $\theta = 90^\circ$, so $n_e(90^\circ) = n_o$, and $\phi = 0$.

The orientational distribution of the director in the cell in the presence of an applied electric field is determined by the strength of the field, the electric permittivity and elastic constants of the liquid crystal, and most importantly by the properties at the interface between the liquid crystal and the aligning surfaces. If the director at the surface satisfies the *strong anchoring* condition i.e. it is unaffected by the applied electric field, and the surface director is strictly perpendicular to the electric field, then the reorientation of the director exhibits a threshold response, known as a Fréedericksz transition. The change in transmitted intensity as a function of voltage can be calculated using continuum theory and simple optics [38.36], and a typical transmission curve for an ECB cell is illustrated schematically in Fig. 38.17.

The threshold voltage for director reorientation is independent of cell thickness, and is given by

$$V_{th} = \pi \sqrt{\frac{K_1}{\epsilon_0 \Delta \epsilon}}, \quad (38.20)$$

where $\Delta \epsilon$ is the anisotropy in the dielectric permittivity and K_1 is the splay elastic constant. However, it is clear from (38.17) and (38.18) that the intensity of light transmitted by the ECB cell depends on the cell thickness and the wavelength of light. This undermines the usefulness of displays based on the ECB mode, since they require cells of uniform thickness and also they will tend to show colouration in white light. Another important operating characteristic of displays is the *angle of view*, i.e. how the image contrast changes as the angle of incidence moves away from 90° with respect to the plane of the cell. This is clear from Fig. 38.16, where the optical paths in the distorted state for observation to the left or right of the perpendicular to the electrodes are clearly different.

In order to construct a useful display from a simple on/off shutter, it is necessary to consider how image data will be transferred to the display. This is known as *addressing*, and to a large extent it is determined by the circuitry that drives the display. However, we shall see that certain properties of liquid crystals also contribute to the effectiveness of different types of addressing. The simplest method of displaying images on a liquid-crystal display is to form an array of separate cells of the type illustrated by the ECB cell, each having a separate connection for the application of an electric field. Images can then be created by switching on, or off,

those cells required to form the image. This technique is known as *direct addressing*, and can be illustrated by the seven-segment displays used in watches and numerical instrument displays (Fig. 38.18).

To create complex displays, a large number of separate cells, known as picture elements or pixels, must be fabricated, usually in the form of a matrix. Providing separate electrical connections for these pixel arrays (e.g. 640×480 for a standard visual graphics array (VGA) computer screen, is impossible, and so other methods of addressing have had to be developed. Historically, the first was the technique known as *passive matrix addressing*, in which the array of cells are identified in rows and columns, and connections are only made to the rows and columns: for an array of $n \times m$ pixels, only $n + m$ connections are made instead of $n \times m$, as required for direct addressing. Each row is activated in turn (sequentially), and appropriate voltage pulses applied to the columns. Only those pixel elements for which the sum of column and row voltages exceeds a *threshold* are switched to an on-state. However, the problem with this method is that many unwanted pixels in the off-state still have a voltage applied, and may be partially activated in the display: so-called *crosstalk*. The time-sharing of activating signals is known as *multiplexing*, and it relies on the addressed pixels holding their signal while other pixels which make up the image are activated. Provided that the multiplexing is on a time scale of milliseconds, any fluctuation in the image goes unnoticed. However, if the multiplexing becomes too slow, the image starts to flicker. In fact there is a limit to the number of rows of the matrix which can

be addressed (n_{\max}), which is related to the ratio of the on-voltage to off-voltage by a result due to *Alt* and *Pleshko* [38.37]:

$$\frac{V_{\text{on}}}{V_{\text{off}}} = \sqrt{\frac{n_{\max}^{1/2} + 1}{n_{\max}^{1/2} - 1}}. \quad (38.21)$$

The cell characteristic which determines the values of V_{on} and V_{off} is the optical transmittance curve, as illustrated in Fig. 38.17. Depending on the desired contrast ratio $I_{\text{on}}/I_{\text{off}}$ for a pixel, then the voltages $V_{\text{on}}/V_{\text{off}}$ are determined. Thus, in the operation of a passively addressed matrix display, there is a tradeoff between the contrast ratio ($I_{\text{on}}/I_{\text{off}}$) and the number of rows i. e. the complexity of the display. The shape of the transmittance curve is determined by the liquid-crystal material properties, and so these affect the resolution of the image displayed. For a desired contrast ratio of 4, the corresponding on/off ratio for an ECB cell might be 1.82. This gives a maximum number of rows as four, which corresponds to a very-low-resolution display, which would be unusable except for a very basic device. This brief description of a simple electro-optic display operating in the ECB mode illustrates that its performance depends on the dielectric, optical and elastic properties of the liquid crystal material used. Additionally, the properties at the liquid-crystal substrate interface and the geometry of the cell will influence the electro-optic response of the cell.

One performance characteristic that is of very great importance is the speed with which the display information can be changed, since this determines the quality and resolution of moving images. The dynamics of fluids are related to their viscosity, and it has already been shown that the viscous properties of liquid crystals are complicated to describe, and are correspondingly difficult to measure. The complication arises because liquid crystals are elastic fluids, and so there is a coupling between the flow of the fluid and the orientation of the director within the fluid. We have seen that the motion of the director can be described in terms of a rotational viscosity, and the optical properties exploited in displays are related to changes in the director orientation. Thus it is the rotational viscosity that is of primary importance in determining the time response of displays. However, the fact that changes of director orientation cause fluid flow in liquid crystals complicates the process.

The time response of a liquid-crystal display pixel can be illustrated by reference to the ECB display, although other cell configurations modify the behaviour to some extent. In what follows, we shall assume that

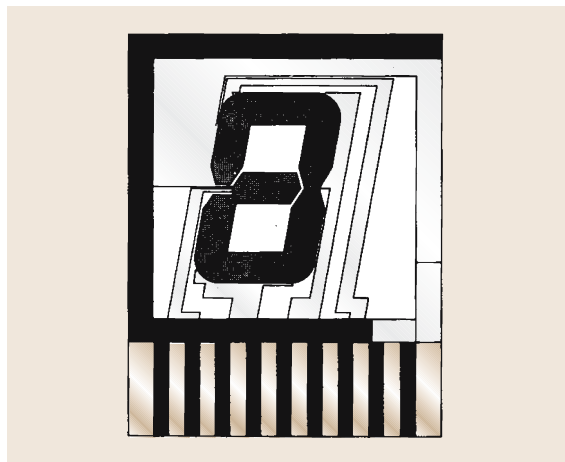


Fig. 38.18 Schematic of a directly addressed seven-segment liquid-crystal display

the reorientation of the director within a display pixel does not cause the nematic liquid crystal to flow. For a uniform parallel-aligned nematic-liquid-crystal film, the time for the director to respond depends on the magnitude of the electric field (or voltage) applied to the cell. If the voltage applied is only just greater than the threshold voltage, then the time is very long, while if a large voltage is applied, then the director responds quickly. It is found that the time response can usually be represented as an exponential behaviour, although effects of flow will change this. Neglecting these, a response time τ_{on} can be defined in terms of the change in transmitted light intensity as

$$\frac{I(t) - I(0)}{I_{\text{on}} - I(0)} = 1 - \exp \left[- \left(\frac{t}{\tau_{\text{on}}} \right) \right]. \quad (38.22)$$

and the relaxation time for switching on the display is given approximately by [38.38, 39]

$$\tau_{\text{on}} = \frac{\gamma_1 \ell^2}{\pi^2 K_{\text{eff}}} \left[\left(\frac{V_{\text{on}}}{V_{\text{th}}} \right)^2 - 1 \right]^{-1}. \quad (38.23)$$

On removing the applied voltage, the display element returns to its off-state, but with a different relaxation time which is independent of the applied voltage, such that

$$\tau_{\text{off}} = \frac{\gamma_1 \ell^2}{\pi^2 K_{\text{eff}}}; \quad (38.24)$$

here γ_1 is the rotational viscosity coefficient, and the effective elastic constant, K_{eff} that appears in these expressions is the splay elastic constant, K_1 , for the ECB-mode display. These equations can be modified for other display configurations by changing K_{eff} .

Although the ECB-mode display is the simplest that can be envisaged, based on the Fréedericksz effect, there are many disadvantages, in particular with respect to its viewing characteristics, and it has not been used commercially to any significant extent. However, the apparently simple modification of twisting the upper plate by 90° has resulted in the phenomenally successful twisted nematic display, which represents a large part of today's multi-billion-dollar market.

38.3.2 High-Volume Commercial Displays: The Twisted Nematic (TN) and Super-Twisted Nematic (STN) Displays

The simple twisted nematic display is essentially the same as the ECB display depicted in Fig. 38.16, except

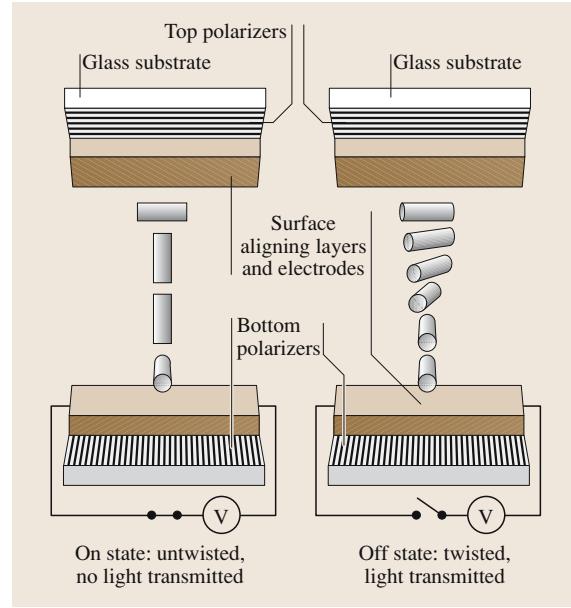


Fig. 38.19 Schematic of a twisted nematic display

that the orientations of the surface director of the containing glass plates are rotated by 90° . However, the TN cell, invented by *Schadt* and *Helfrich* in 1970 [38.40], represented a considerable improvement over earlier devices, and rapidly became the preferred configuration for commercial displays. A schematic representation of the TN cell is given in Fig. 38.19.

This twisted configuration for a liquid crystal film was discovered by *Mauguin*, who found that instead of producing elliptically polarised light, such a twisted film could rotate the plane of polarisation by an angle equal to the twist angle between the surface directors of the glass plates [38.41]. In fact by working through the optics of twisted films, *Mauguin* showed that perfect rotation of the plane of polarisation only resulted if the film satisfied the following condition:

$$2\Delta n \ell \gg \lambda. \quad (38.25)$$

Thus for the cell illustrated in Fig. 38.19, the off-state would be perfectly transmitting. This is known as the *normally white* mode. In reality the cells used for TN displays do not meet the *Mauguin* condition, and the transmission for a 90° twisted cell between crossed polarisers and incident unpolarised light is given by [38.42]

$$T = \frac{1}{2} \left(1 - \frac{\sin^2 \left(\frac{\pi}{2} \sqrt{1 + u^2} \right)}{1 + u^2} \right), \quad (38.26)$$

where $u = \frac{2\Delta n \ell}{\lambda}$. Equation (38.26) shows that, for sufficiently large u , the transmission, T is indeed a maximum of 0.5, however it is also a maximum for $u = \sqrt{3}, \sqrt{15}, \sqrt{35}$, etc. These points on the transmission curve correspond to the *Gooch–Tarry minima*; they are labelled as minima, since they were found for a TN cell operating in the *normally black* state. Most commercial cells operate under conditions of the first or second minima so that thin cells can be used, which give faster responses. It is, therefore, important that the birefringence of the liquid-crystal material can be adjusted to match the desired cell thickness, so that the display can have the best optical characteristics in the off-state.

Application of a sufficiently strong electric field across the twisted film of a nematic liquid crystal having a positive dielectric anisotropy causes the director to align along the field direction. Under these circumstances the film no longer rotates the plane of polarised light, and so appears dark. The transmission as a function of voltage for a twisted cell is similar to that shown in Fig. 38.17, except that the transmission varies more strongly with change in voltage above the threshold, and drops to zero much more rapidly. In contrast to the ECB cell discussed above, the threshold voltage for a TN cell depends on all three elastic constants:

$$V_{\text{th}}^{\text{TN}} = \pi \sqrt{\frac{\left(K_1 + \frac{\zeta}{2}(K_3 - 2K_2)\right)}{\varepsilon_0 \Delta \varepsilon}}, \quad (38.27)$$

where ζ is the twist angle (usually $\pi/2$). The relative change of the transmission intensity with voltage of the TN cell is greater than for the ECB cell, and it can be shown that the steepness of the transmission curve increases as the property ratios K_3/K_1 and $\Delta \varepsilon/\varepsilon_{\perp}$ decrease. The on/off voltage ratio for a TN cell is closer to unity, than for an ECB cell, and so for similar contrast ratios more lines can be addressed: up to about 20 for typical cells and materials. This is significantly larger than for the ECB cell, and so the TN cell allows more complex images to be displayed. There is still a wavelength dependence for the transmission, although this is less marked than for the ECB mode. However, even with the improved multiplexing capabilities of the TN display over the ECB cell, it is still not good enough to use for computer screens. One very successful approach to solve this problem was to modify the TN cell geometry so that instead of a 90° twist, the directors on opposite sides of the cell are rotated by about 270° . This is known as a super-twisted nematic cell.

The concept of the 270° super-twisted nematic (STN) display seems at first sight to be irrational

[38.43, 44]. The director is an apolar vector and so there should be no difference between a 270° and a 90° twisted cell. However, it is possible to maintain a director twist greater than 90° if the liquid crystal is chiral. The use of chiral additives in 90° TN cells was already established, since a very small quantity of chiral dopant added to a TN mixture would break the left/right twist degeneracy in the cell and so remove patches of reversed twist, giving a much improved appearance to the display. If the amount of chiral dopant was increased, then the chiral liquid-crystal mixture would develop a significant intrinsic pitch. By adjusting the concentration of the chiral dopant, the pitch of the mixture could be matched to the 270° twist across the cell thickness of ℓ . i. e. $p \approx 4\ell/3$. An STN cell operates in the same way as a TN cell, so that an applied electric field causes the director to rotate towards the field direction, thereby changing the optical retardation through the film and the transmission between external crossed polarisers. However, the additional twist in the STN cell has a significant effect on the optical properties of the nematic film.

The optical characteristics of the STN off-state are usually outside the Mauguin condition, which means that polarised light passing through the cell is not guided, and emerges elliptically polarised. The degree of ellipticity is wavelength dependent, and so in white light the off-state appears coloured, as does the on-state. An ingenious solution to this problem is to have two identical STN cells, one behind the other, but where the second compensating cell has a twist of the opposite sense. In operation, only the first of the cells has a voltage applied to it. The compensation cell acts to subtract the residual birefringence of the liquid crystal layer, and the display now switches between white and black. Despite this additional complexity, the huge advantage of the STN display is the rapid change in optical transmission with increasing voltage, and a full optical analysis shows that under optimum conditions the rate of change of transmission with voltage can become infinite. In modern implementations of the STN display, the residual birefringence can be compensated by an optical film, avoiding the need for double cells. It is easy to operate an STN display with a $V_{\text{on}}/V_{\text{off}}$ ratio of 1.1, which corresponds to an n_{max} of 100. The shapes of the transmission/voltage curve of ECB, TN and STN cells are a direct consequence of the dielectric and elastic properties of the liquid crystal material, but also depend strongly on the configuration of the cells and the surface alignment

In the description given of displays and their performance, some important aspects have been ignored. The

surface orientation of the directors has been assumed to be pinned in the surface plane, which is the requirement for a threshold response. However, it has been found that the performance of displays can be greatly improved if this condition is relaxed, and a pre-tilt is introduced to the cell, such that the surface director may make an angle of up to 60° to the plane of the containing glass plates. This pre-tilt can be introduced by different surface treatments, and it depends on the interfacial properties of the liquid crystal. An important performance characteristic of displays is the angle dependence of the contrast ratio, or more simply how the appearance of the displayed image changes with angle of view. This is largely determined by the display design, and can be accurately calculated from the optical properties of the cell. The refractive indices of the liquid crystal will affect the contrast ratio and angle of view, but precise control of these performance parameters is difficult. Improvements to the appearance in terms of the angle of view or brightness of displays have been achieved by the placing precisely manufactured birefringent polymer films behind or in front of liquid-crystal cells.

38.3.3 Complex LC Displays and Other Cell Configurations

The STN configuration described above used passive matrix addressing, and this opened up the possibility of relatively large-area high-resolution displays, which could be used in laptop computers and other hand-held displays. The next step was to introduce colour by di-

viding each picture element into three sub-pixels with red, green and blue filters. However, the demand in some market sectors for larger displays with improved appearance having higher resolution [extended graphics array (XGA) displays have 1024×768 pixels] overwhelmed the capability of passive matrix addressing, and the alternative method of active matrix addressing is now used for more complex displays. This technique requires each pixel to have its own switch, as in the simple seven-segment display already described. For high-resolution displays hundreds of thousands of switches are provided by thin-film transistors deposited onto the glass substrate, which forms the screen. These are known as thin-film-transistor twisted-nematic (TFT-TN) displays [38.45]. The sophistication of these displays relies on the capabilities of integrated circuit technology, but the properties of the liquid-crystal materials must still be optimised for the device configuration. A representation of a TFT-TN display is given in Fig. 38.20.

Although being described as an active matrix, the TFT-TN display still uses sequential addressing of pixel rows, and so activated pixels must remain switched on while the rest of the display image is created. The TFTs provide a source of voltage to each liquid crystal pixel, which must then hold its charge, as a capacitor, until the image is changed. So another property of liquid crystals becomes important, that of low electrical conductivity, since the charge on a pixel will be lost by conduction through the liquid crystal. This determines the choice of materials for TFT-TN displays. Generally, high dielectric anisotropy is a desirable property for liquid-crystal display mixtures, since it reduces the operating voltage. However, materials with a high dielectric constant tend to have a high electrical conductivity, since charges either from impurities or leached from surfaces will be stabilised in high-dielectric-constant fluids. Thus the selection of suitable materials requires a compromise between its dielectric and conductance properties, and, of course, the all important refractive indices.

Many different cell configurations, which exploit the optical properties of liquid crystals in different ways, have been tried, and some of these have been commercialised to meet particular market requirements. One rather successful approach to the problem of restricted viewing angle has been the development of the in-plane switching mode (IPS) twisted nematic display [38.46, 47]; this is illustrated in Fig. 38.21. The two optical states of the cells are (i) twisted, and (ii) planar (parallel aligned film), and the director is switched between these states by application of an electric field across electrodes on a single plate of the cell. The state

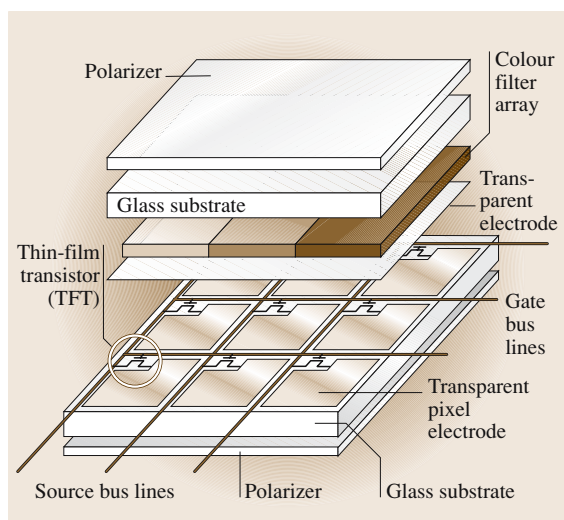


Fig. 38.20 Schematic of a complex colour TFT-TN display

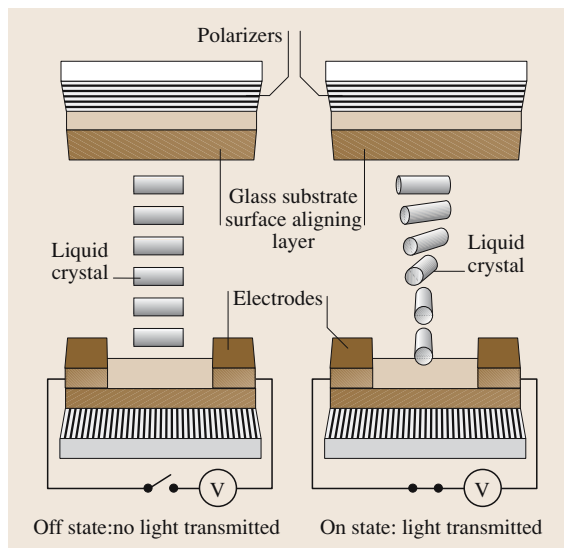


Fig. 38.21 Schematic of an in-plane switching mode display

that is stabilised by the electric field depends on the dielectric anisotropy of the liquid-crystal material. The preferred configuration uses materials having a negative dielectric anisotropy, so that the off-state is a planar-aligned liquid-crystal film. Application of an electric field to the in-plane electrodes will cause the director at the bottom surface to align perpendicularly to its initial direction, and so induce a twist through the cell (Fig. 38.21).

From an optical point of view, the director is always in the plane of the cell, and this means there is less distortion of an image when viewed at angles other than 90° . Another advantage of the IPS device is that the electric field is confined to the lower plate, and the lines of force do not extend across the cell to the grounded upper plate. This means that a very low electrical conductivity of the liquid-crystal material is less important than for conventional TFT-TN displays. The threshold voltage for the IPS-mode device is given by

$$V_{\text{th}}^{\text{IPS}} = \frac{\pi d}{\ell} \sqrt{\frac{K_2}{\varepsilon_0 \Delta \varepsilon}}, \quad (38.28)$$

where ℓ is the thickness of the liquid-crystal film, and d is the separation of the in-plane electrodes. Not surprisingly the threshold depends only on the twist elastic constant, which is usually smaller by about a factor of two than the splay and bend elastic constants. While this helps to reduce the operating voltage, the smaller elastic

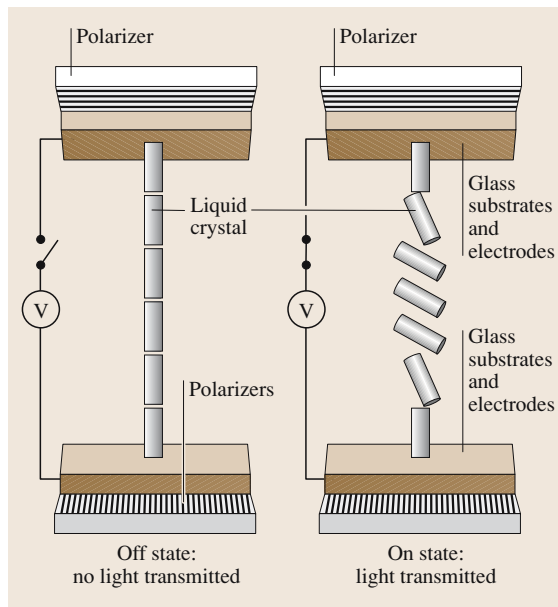


Fig. 38.22 Schematic of a twisted vertically aligned display

energy associated with the pure twist deformation results in longer switching times. A further disadvantage of the IPS display is that the optical transmission of the cell is reduced by the requirement to have both electrodes deposited on one plate, thereby making smaller the active area available to display the image.

A configuration which shares some of the characteristics of the IPS cell and has been successfully commercialised is the twisted vertically aligned (TVAN) cell. The two optical states for this configuration are uniform vertical (homeotropic) alignment of the director for the off-state, and a twisted geometry for the on-state (Fig. 38.22). The liquid crystal material used has a negative dielectric anisotropy, so application of an electric field between the plates causes the director to align perpendicularly with respect to the field direction. A small quantity of optically active material (chiral dopant) is added to the liquid-crystal mixture to ensure that the switched director adopts a twisted configuration through the cell. The advantages of this cell are good viewing-angle characteristics and high optical contrast. Improvements in display technology continue to be made, often simplifying earlier devices. For example, high-quality displays described as vertically aligned nematic (VAN) devices are now available based on the TVAN configuration, but without the twist. The material requirement here is for a liquid crystal of negative dielectric anisotropy,

that will align perpendicularly to an applied electric field.

In the devices described above, one state is defined by the surface conditions of the cell, while the other is defined by the action of the applied electric field. A bistable device is one in which two stable field-free states exist, both of which are accessible by switching with an external field. The first bistable liquid-crystal display to be developed was based on a ferroelectric effect observed in chiral tilted smectic C liquid crystals [38.48]. This ferroelectric smectic display has achieved some limited commercial success in specialist markets, but relies on a surface stabilisation of smectic layers, which is very sensitive to mechanical shock. Recently [38.49], bistable nematic displays have been developed in which two alignment states within a liquid-crystal cell, having different optical transmission, can be stabilised. If one of the substrates of a normal cell is replaced by a surface which has potentially two states of minimum energy corresponding to two surface alignments of the director, then it becomes possible to switch these states selectively using an electric field. A suitable bistable surface is provided by a grooved surface (grating) which has been treated with a surfactant to favour homeotropic alignment of the director at its surface [38.50]. Thus the two surface states correspond to (i) that determined by the grating, and (ii) that determined by the surfactant where the director is perpendicular to the substrate. Combining this intrinsically bistable substrate with a second substrate having a director alignment direction perpendicular to the grating direction gives a cell configuration capable of supporting two optically distinct stable states, which can be switched between using an applied voltage. Various director configurations are possible with this type of cell, and one example is illustrated in Fig. 38.23.

In the absence of any perturbation, the director orientation within the cell will be determined by the homeotropic alignment at one substrate and the alignment at the grating substrate. This hybrid (uniform planar and homeotropic) alignment causes a spatially varying director tilt through the sample. Application of an electric field to a positive-dielectric-anisotropy material will cause the director to align parallel to the applied field, and eventually a fully homeotropic configuration for the director is stabilised. The switch back from the homeotropic state to the hybrid state is thought to be due to a flexoelectric interaction. Other alignment configurations are also possible for the so-called zenithal bistable nematic (ZBD) cell. Displays based on these configurations share the good viewing characteristics of both the IPS and TVAN configurations, but they have the great

advantage that the image is retained when the voltage is removed.

All the displays described so far rely on the coupling between an applied electric field and the dielectric properties of the liquid-crystal material, but, as we have shown, other material properties are just as important to the operating characteristics of the display. The appearance of a display depends on the optical properties of the liquid crystal and the cell configuration, but the operating voltage and switching times of a display are crucial in determining the types of application. Changing the nature of the interaction between the switching electric field and the liquid crystal gives rise to another range of possibilities for liquid-crystal devices. Under certain circumstances, a liquid crystal can be made to exhibit permanent ferroelectric (or spontaneous) polarisation, and this couples linearly with an external electric field, in contrast to dielectric properties which couple with the square of the electric field strength. Not surprisingly this makes a big difference to the switching behaviour of liquid crystals.

The final display to be considered is based on flexoelectric coupling between the electric field and the liquid crystal. Flexoelectricity occurs, in principle, with all liquid crystals, chiral or not, and shows itself as a bulk electric polarisation induced by an elastic strain. Conversely application of an electric field can cause an elastic strain. In general, flexoelectricity is rather small and difficult to detect, however it is thought to be responsible for an electro-optic effect observed in chiral nematic liquid crystals, which is being investigated for display applications. The effect, sometimes known as the deformed helix mode [38.51], is similar in some respects to the ferroelectric switching observed in chiral

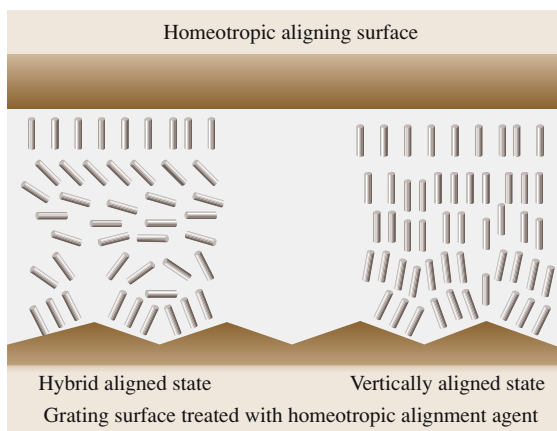


Fig. 38.23 Schematic of a zenithal bistable device

smectic C phases, but there is no longer a requirement for a layered structure. Chiral nematic liquid crystals spontaneously form helical structures in which the director rotates with a pitch determined by the molecular structure. If an electric field is applied perpendicularly to a chiral nematic helix, then there is a tendency for the helix to unwind, depending on the sign of the dielectric anisotropy. Even if the dielectric anisotropy of the material is zero, there is an elastic strain which can generate a polarisation (flexoelectric polarisation), which will interact with an applied electric field. This may be exploited in a device configuration, where a thin film of a chiral nematic liquid crystal, having a small or zero dielectric anisotropy, is aligned such that its helix axis is parallel to the containing glass plates (Fig. 38.24).

Application of an electric field across the plates will cause a distortion of the helix through the splay and bend flexoelectric coefficients, which appears as a rotation of the optic axis in the plane of the film [38.25]. Reversal of the electric field direction will reverse the rotation of the optic axis, with an intrinsic switching time about one hundred times faster than conventional nematic displays. Optically, the effect observed is very similar to that exhibited by smectic ferroelectric displays.

There are many cell configurations that can be used with liquid crystals to produce optical switches, displays or light modulators, and some of the more important have been described. The precise operation and performance of these liquid-crystal devices depends on both the cell design and the material properties of the liquid crystal. To a large extent the configuration of the liquid crystal within the cell is determined by such factors as the surface treatment of the plates enclosing the liquid crystal and the interactions between the surfaces and the liquid crystal. Our understanding of these interactions is very limited at the present time, and much more research is necessary before a quantitative theory can be formulated. However, given the cell configuration, the performance of the liquid-crystal device depends crit-

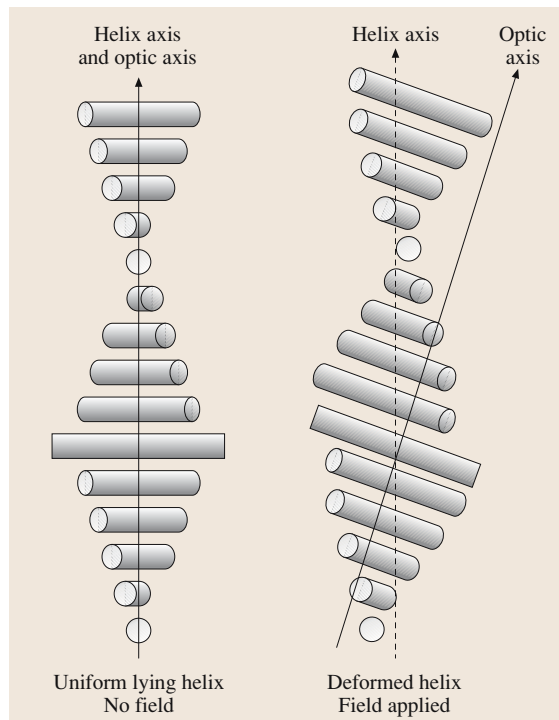


Fig. 38.24 Schematic of a deformed helix mode flexoelectric display

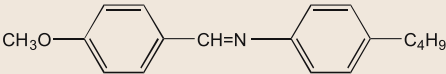
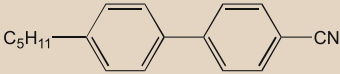
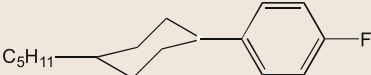
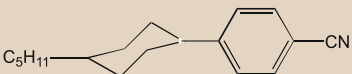
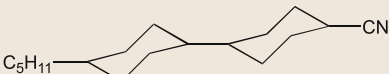
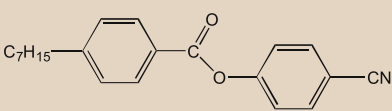
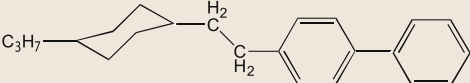
ically on the physical properties of the liquid-crystal material. Thus the electrical switching characteristics will depend on the dielectric properties, while the optical appearance of the device will be determined by the refractive indices. Elastic properties contribute to both the electric field response and the optical appearance, since any deformation of the director will be determined by the elastic properties of the liquid crystal. Finally, the all-important dynamical behaviour will be controlled by the viscous properties of the liquid crystal. All these material properties will be discussed in the next section.

38.4 Materials for Displays

The most important requirement for a liquid-crystal display material is that it should be liquid crystalline over the temperature range of operation of the device. Despite this, some of the first experimental display devices incorporated heaters in order to maintain the material in the liquid-crystal phase: for example, the first liquid-crystal shutter patented by Marconi and

early prototype displays developed by the Radio Corporation America (RCA). It was not until the late 1960s that room-temperature nematic liquid crystals suitable for display applications were discovered. The first of these were based on Schiff's bases, which although easy to prepare, were difficult to purify and were susceptible to chemical decomposition in a de-

Table 38.1 Some typical liquid-crystal materials, including selected physical properties (ex – extrapolated from measurements on a nematic solution)

Compound Transition temperatures (°C) [ref.]	$\Delta\epsilon$ (T °C)	Δn (T °C)	Rotational Viscosity (γ_1 /m Pa s)
 T_{CrN} 22 °C T_{NI} 47 °C [38.53]	≈ 1	0.2	$\gamma_1 = 109$ [38.52] (37 °C)
 T_{CrN} 22 °C T_{NI} 35 °C [38.53]	8.5 (29 °C)	0.18 (25 °C)	$\gamma_1 = 102$ [38.52] (25 °C)
 T_{CrI} 34 °C [38.54]	3.2(ex)	0.05(ex)	
 T_{CrN} 31 °C T_{NI} 55 °C [38.53]	9.9 (48 °C)	0.12 (40 °C)	$\gamma_1 = 128$ [38.52]
 T_{CrN} 62 °C T_{NI} 85 °C [38.53]	3.5 (78 °C)	0.05	
 T_{CrN} 44 °C T_{NI} 57 °C [38.53]	19.9 (40 °C)	0.15	
 T_{CrN} 67 °C T_{NI} 82 °C [38.55]	≈ 0	0.15	

vice. One material which attracted particular attention from experimental physicists was MBBA, which has a nematic range of 22–47 °C (Table 38.1). The basic two-ring core linked by an imine group was the structural unit of many compounds having different terminal groups which were prepared for display mixtures in the early 1970s. It was found that mixtures of Schiff's bases often had lower crystal-to-nematic transition temperatures than any of the components, and furthermore would often remain liquid crystalline even below the thermodynamic crystallisation temperature. These two phenomena of eutectic behaviour and

super-cooling have been exploited in the development of materials for devices. The early experiments on liquid-crystal displays were primarily focused on nematic or chiral compounds, but we have seen that other displays have been developed which use different liquid-crystal phases, most importantly the chiral smectic C phase.

Early studies established guiding principles for the development of display materials. First, the phase behaviour must be acceptable, i.e. the right phase stable over a suitable temperature range. Secondly, the material must have the correct electrical and optical proper-

ties for the particular display application envisaged, and above all must be of sufficient chemical purity to prevent any deterioration in performance over time. Again, guided by the early experiments, suitable display materials require the synthesis of compounds of appropriate chemical structure, and then the formulation of mixtures to optimise the properties. There have been a number of reviews of liquid-crystal materials for displays [38.53, 54, 56–59] and these contain many tables of data on a wide range of compounds. In this Section, we will give a brief account of the basic chemical structures used for materials in modern liquid-crystal displays, and then show how mixtures are devised to give the best possible performance characteristics for different displays. It has to be recognised that many of the details of display materials are matters of commercial confidentiality, and so it is not possible to give precise accounts of materials currently used or under investigation. However, the generic chemical structures and principles used in developing suitable mixtures are generally applicable.

38.4.1 Chemical Structure and Liquid-Crystal Phase Behaviour

There is a huge literature on the relationship between the structure of mesogens and the nature and stability of the liquid-crystal phases they form [38.60]. The studies have embraced empirical correlations of chemical structure and phase behaviour, theoretical calculations for simple particles (hard rods, spherocylinders etc.) representing mesogens, and computer simulations of collections of particles of varying complexities. For the display applications considered in this Chapter, the desired phases are nematic, and occasionally chiral nematic or chiral smectic C. Such phases are formed by molecules having extended structures, which usually require the presence of terminal alkyl chains to reduce the crystallisation temperatures. Components in nematic display mixtures typically have two, three or four carbocyclic rings joined directly or through a variety of linking groups.

38.4.2 The Formulation of Liquid-Crystal Display Mixtures

The two requirements for a liquid crystal to be used in a display are a suitable temperature range of phase stability and appropriate physical properties. These requirements cannot be satisfied for complex displays by a single compound, and commercial display materials may contain up to twenty different components. The

formulation of these mixtures is essentially an empirical process, but guided by the results of thermodynamics and experience. The principles behind the preparation of multicomponent mixtures can be illustrated initially by consideration of a binary mixture.

It is well-known that the melting point of a binary mixture of miscible compounds is depressed, sometimes below the melting points of both components. Furthermore, the melting point of the binary mixture may exhibit a minimum at a particular composition, known as the eutectic. This occurs with liquid-crystalline compounds, and provides a method of reducing the lower temperature limit for liquid-crystal phase stability in mixtures. The upper temperature limit of the liquid-crystal range is fixed by the transition to an isotropic liquid. The phase rule of Willard Gibbs predicts that in binary mixtures there will always be a region of two-phase coexistence in the vicinity of a phase transition; that is, the transition from liquid crystal to isotropic occurs over a range of temperatures for which both the isotropic liquid and liquid crystal are stable in the mixture. Because of the weak first-order nature of most liquid crystal to isotropic phase transitions, the two-phase region is small. The character of phase transitions is determined by the corresponding entropy change, and a weak first-order transition has a small ($\approx 2 \text{ J K}^{-1} \text{ mol}^{-1}$) associated entropy. If the latter were zero, then the transition would be second order, and there would no longer be a region of two-phase coexistence. The phase diagrams of multicomponent nematic mixtures can be calculated by thermodynamic methods [38.61, 62] and the transition temperatures of the mixtures can vary with composition in a variety of ways. For mixtures of two liquid-crystalline compounds of similar chemical constitution, the variation of the nematic to isotropic transition temperature is approximately linear with composition [38.63].

It is possible to calculate the variation of the melting point with composition using an equation attributed to Schroder and van Laar. For each component i , the mixture composition (mole fraction x_i) and the melting point of the mixture T are related by

$$\ln x_i = -\frac{\Delta H_i}{R} \left(\frac{1}{T} - \frac{1}{T_i} \right), \quad (38.29)$$

where ΔH_i and T_i are, respectively, the latent heat of fusion and melting point of the pure component i . For a binary mixture there are two such equations which can be solved to give the eutectic temperature and composition. In a multicomponent mixture, the set of

equations (38.29) can be solved subject to the condition,

$$\sum_i x_i = 1 \quad (38.30)$$

to predict the eutectic of the mixture.

While there is a reasonable thermodynamic basis to the prediction of the phase diagrams of mixtures, the determination of the physical properties of mixtures from the properties of individual components is much more difficult. Given the absence of any better theories, it is common to assume that in mixtures, physical properties such as dielectric anisotropy, birefringence and even viscosity vary linearly with the amount of any component, at least for small concentrations. While this may give an indication of the effect of different components on the properties of a display mixture, it can also be very misleading. One theoretical problem is that, for a mixture at a particular temperature, the orientational order parameters of the different components are not equal. The more anisometric components (e.g. three-ring mesogens) are likely to have a larger orientational order parameter than smaller (two-ring) mesogens. Since the various physical properties of interest in displays depend on the order parameter in different ways, it is difficult to predict the contribution of different components to the overall mixture properties. Despite this, many tables of data for liquid-crystal compounds of interest for display mixtures are prepared [38.60] on the basis of extrapolated measurements on mixtures at low composition, normally < 20% w/w. There is always a problem concerning the temperatures at which to compare the physical properties of liquid crystals and their components. Many measurements are made at room temperature, so that this becomes the temperature for comparison. However, a more useful approach is to compare properties at equal reduced temperatures (or at the same shifted temperature, $T_{NI} - T$), since under these conditions the orientational order parameters are likely to be similar.

38.4.3 Relationships Between Physical Properties and Chemical Structures of Mesogens

Electrical and Optical Properties

These properties include the dielectric permittivity, electrical conductivity and refractive indices. The magnitude of the dielectric anisotropy determines the threshold voltage necessary to switch a display, and influences the transmission/voltage characteristics of the cell. Depending on the particular display configuration, a positive or

negative dielectric anisotropy may be required. Refractive indices strongly affect the appearance of a display. Usually the refractive indices or birefringence must be adjusted for a particular cell configuration to give the optimum on/off contrast ratio. Coloration in displays can sometimes occur in materials of high refractive index, and so it is desirable to keep the birefringence as low as possible, compatible with an acceptable contrast ratio. For twisted structures, the magnitude of the birefringence also determines the efficiency of light guiding, and so close control of the values of the principal refractive indices of a display mixture is important. For non-conducting materials, the refractive indices are measures of the dielectric response of a material at very high i.e. optical frequencies, and it is possible to formulate a single theory which relates the dielectric and optical properties of a liquid crystal to its molecular properties. Unfortunately this is not possible for the electrical conductivity. The latter is largely determined by the purity of the liquid crystal, but it is found that the higher the value of the permittivity, the larger the electrical conductivity. Materials of high electrical conductivity tend to leak charge, and so an image may deteriorate during a multiplexing cycle. In general it is desirable to minimise the conductivity of a display mixture, although this was not the case for the first liquid-crystal displays reported [38.64]. These utilised the strong light scattering which results when an electric field is applied to certain nematic materials. The scattering is due to electrohydrodynamic instabilities in liquid-crystal materials which have a significant electrical conductivity. Such materials are not suitable for use in modern, fast-multiplexed displays.

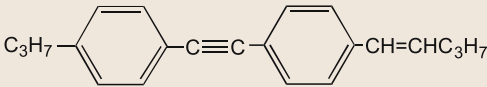
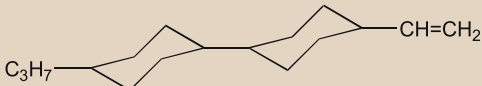
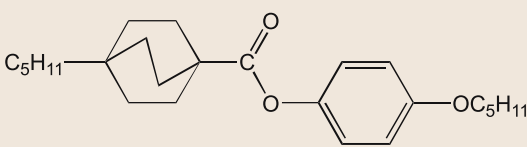
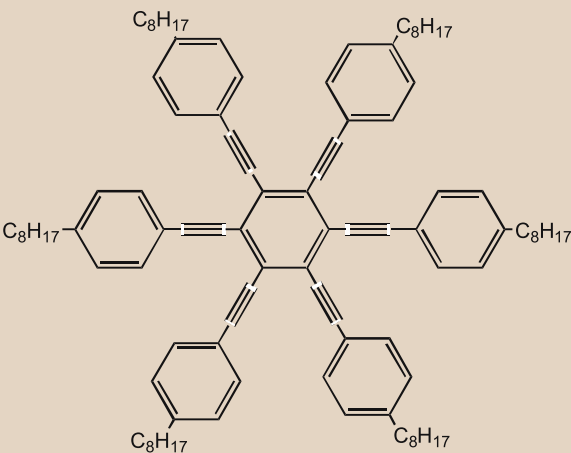
The dielectric anisotropy, $\Delta\epsilon$, and birefringence, Δn , of a nematic can be related to molecular properties of polarisability and dipole moment using a theory originally developed by *Maier* and *Meier* [38.65]. The birefringence is given by

$$\Delta n \approx \frac{NS}{\epsilon_0} (\alpha_\ell - \alpha_t), \quad (38.31)$$

where N is the density in molecules per m^3 and $\Delta\alpha = (\alpha_\ell - \alpha_t)$ is the anisotropy of the molecular polarisability. S is the order parameter, defined in Sect. 38.2.1, and small corrections due to the local field anisotropy have been neglected. Such corrections cannot be ignored in the corresponding expression

$$\Delta\epsilon = \frac{NhFS}{\epsilon_0} \left[\Delta\alpha + \frac{\mu^2}{2k_B T} (3 \cos^2 \beta - 1) \right] \quad (38.32)$$

Table 38.2 Materials of high, low and negative birefringence

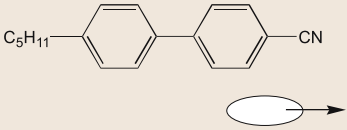
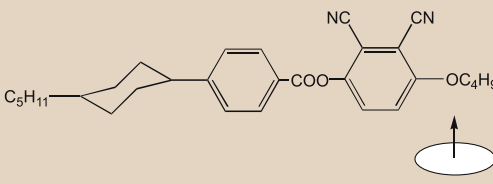
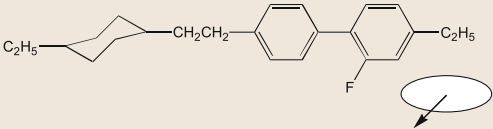
Compound	Transition temperatures	Δn [ref.] ($T^{\circ}\text{C}$)
	T_{NI} 112 °C	0.31 [38.66] (92 °C)
	T_{CrSmB} 23 °C T_{SmBN} 35 °C T_{NI} 49 °C	0.052 [38.59]
	T_{NI} 93.5 °C	0.074 [38.66] (58.5 °C)
	T_{CrN} 80 °C T_{NI} 96 °C	-0.193 [38.66] (61 °C)

for the dielectric anisotropy, especially for materials of high permittivity. In (38.32) h and F are local-field correction factors, while μ is the molecular dipole moment, and β is the angle between the dipole direction and the long axis of the molecule. For molecules containing a number of dipolar groups, μ is the root mean square of the vector sum of all contributing groups. Both the birefringence and the dielectric anisotropy increase with decreasing temperature, and the detailed variation with temperature is largely determined by the temperature dependence of the order parameter S .

The manipulation of birefringence is achieved by changing the chemical constitution of the mesogen. Thus extending the electronic conjugation along the axis of a mesogen will result in an increase in longitudinal polarisability, and hence an increase in birefringence. Saturated carbocyclic rings, such as cyclohexane, and

aliphatic chains generally have small polarisabilities and mesogens containing a predominance of such moieties form low-birefringence liquid crystals. Mixtures for displays require a positive birefringence, which is associated with calamitic or rod-like mesogens. In order to improve the viewing characteristics of displays, optically retarding films are placed in front of the display, and depending on their function, these may be of negative or positive birefringence. The latter can be fabricated by encapsulation or polymerisation of suitable molecules of an extended structure. On the other hand, films of negative birefringence have been made using discotic materials: i. e. mesogenic molecules formed from disc-like structures which have a negative polarisability anisotropy (Sect. 38.1.3). Some examples of liquid crystals having different birefringence are shown in Table 38.2.

Table 38.3 Materials of high, low and negative dielectric anisotropy; the *inset* figure indicates the direction of the total dipole moment with respect to the core of the molecule ($1 \text{ Debye} = 3.335\,64 \times 10^{-30} \text{ Cm}$)

Compound	Transition temperatures	$\Delta\epsilon$ [ref.] ($T^\circ\text{C}$)	Total dipole moment μ (Debye)
	$T_{\text{CIN}} 22.5^\circ\text{C}$ $T_{\text{NI}} 35^\circ\text{C}$	11.5 [38.67] (25°C)	4.8
	$T_{\text{CIN}} 143^\circ\text{C}$ $T_{\text{NI}} 150^\circ\text{C}$	-10.0 [38.68] (145°C)	6.4
	$T_{\text{CIN}} 13^\circ\text{C}$ $T_{\text{NI}} 64^\circ\text{C}$	0.0 [38.69] (58°C)	1.4

The introduction of chirality into liquid crystals has important consequences for their optical properties. The selective reflection of coloured light from the helical structure of a chiral nematic has already been mentioned in Sect. 38.1.2. All chiral materials will rotate the plane of incident polarised light, and the particular optical properties associated with chiral thin films are exploited in many liquid-crystal device applications.

The dipole moment of a molecule is increased if strongly electronegative or electropositive groups are substituted into the structure, with the result that the dielectric permittivity increases. For mesogenic molecules, the locations of the electropositive or electronegative groups are important, since not only the magnitude but also the orientation of the molecular dipole strongly influences the dielectric anisotropy. From (38.32) it can be seen that the dipolar contribution to the dielectric anisotropy may be positive or negative depending on the value of the angle β , since for values of β greater than 54.7° the dipolar contribution to the permittivity anisotropy becomes negative. This is illustrated by the mesogens shown in Table 38.3, where different structures can be designed to give large positive, negative or zero dielectric anisotropy.

The most effective substituent for producing materials of high dielectric anisotropy is the cyano group, and mixtures containing cyano-mesogens were the basis

for the rapid development of complex displays in the 1980s and early 1990s. However, these mixtures tended to have relatively high viscosities, which gave rise to slow switching times. Another disadvantage was the high electrical conductivity associated with the high dielectric anisotropies which caused charge leakage during multiplexing, and hence degradation of the image.

As the demands placed on liquid-crystal materials by more sophisticated display technologies have increased, new families of molecules have been synthesised and screened for their physical properties. However, it is no longer the properties of the pure mesogens that are of interest, rather how they behave in mixtures. For this reason, the physical properties of most components of display mixtures are measured in mixtures, and values for the *pure* mesogens are obtained by extrapolation. Data derived in this way are useful in designing display mixtures and for comparison purposes, but cannot be relied upon to give quantitative information about the relationship between molecular structure and physical properties.

The major display technologies using TN, STN or TFT-TN configurations require display mixtures having a positive dielectric anisotropy. Many materials have been developed to improve the electro-optical behaviour of these displays, particularly using fluorine-substituted mesogens to provide the required dielectric

and optical properties (for examples see [38.54, 58, 70]). However, within the past seven years, new display configurations have emerged, such as the in-plane switching (IPS) and vertically aligned (VAN and TVAN) nematic modes, which require mixtures with negative dielectric anisotropy. Using the design strategy illustrated above for simple mesogens, it has been possible to prepare a large number of materials with the desired negative dielectric anisotropy. These are again mostly based on fluorine-substituted compounds, and as before their properties have mostly been determined by extrapolation of measurements on mixtures.

Elastic Properties

The property known as elasticity is characteristic of liquid crystals, and distinguishes them from isotropic liquids. It has been shown in Sect. 38.2.3 that the macroscopic orientational disorder of the director in liquid crystals can be represented in terms of three normal modes, designated as splay, twist and bend, and associated with each of these deformations is an elastic constant. Since the elastic properties of display materials contribute to the electro-optic response, their optimisation for particular display configurations is important to maximise the performance of commercial devices. However, despite their importance, the relationships between the magnitudes of elastic constants and the chemical structure of mesogens are poorly understood. There is a good reason for this; the optical and dielectric properties are to a first approximation single particle properties. That is they are roughly proportional to the molecular number density and are also linearly dependent on the order parameter. Because elastic properties are a measure of the change in energy due to displacements of the director, they are related to the orientation-dependent intermolecular forces. Thus, at a molecular level, elastic properties are two-particle properties, and are no longer linearly proportional to the number density. A further consequence is that the elastic properties depend to lowest order on the square of the order parameters. Molecular theories of elasticity in nematic liquid crystals have been developed [38.71] and the simplest results suggest that the different elastic constants can be related to molecular shape

$$K_1 = K_2 \propto \langle x^2 \rangle \text{ and } K_3 \propto \langle z^2 \rangle, \quad (38.33)$$

where $\langle z^2 \rangle$ and $\langle x^2 \rangle$ are the average intermolecular distances parallel and perpendicular to the molecular alignment direction, respectively. Thus theory predicts that for rod-like molecules the splay elastic constant

should be smaller than the bend elastic constant, and increasing the molecular length should increase K_3 , while increasing the molecular width should increase K_1 . This is roughly in accord with experimental results, except that the prediction of equal splay and twist elastic constants is not confirmed (Fig. 38.9). In general, the twist elastic constant is about one half of the splay elastic constant. Hard particle theories [38.72] evaluated for spherocylinders provide further guidance on the relationship of elastic constants to molecular shape. These theoretical results can be presented in a simplified way as follows:

$$\begin{aligned} \frac{K_1 - K}{K} &= \Delta(1 - 3\sigma); \quad \frac{K_2 - K}{K} = -\Delta(2 + \sigma); \\ \frac{K_3 - K}{K} &= \Delta(1 + 4\sigma), \end{aligned} \quad (38.34)$$

where $K = \frac{1}{3}(K_1 + K_2 + K_3)$. The quantities Δ and σ are parameters of the theory, where Δ is approximately equal to the square of the length:width ratio of the spherocylinder, and σ depends on the degree of orientational order. Despite the fact that details of internal chemical structure are ignored, these theoretical results for nematics are in approximate agreement with experimental measurements on simple nematics. If the nematic material has an underlying smectic phase, or if there is a tendency for local smectic-like ordering, this can strongly affect the elastic constants. Both the twist and bend elastic constants are infinite in a smectic phase, and in a nematic phase they diverge as the transition to a smectic phase is approached.

The elastic constants contribute directly to the threshold voltage and the response times of displays. Threshold voltages increase with increasing elastic constants, and the elastic constants responsible depend on the configuration of the display. Thus for the planar-aligned electrically controlled birefringence display (ECB), the threshold voltage depends on K_1 , while the switching voltage for TN displays depends on a combination of all three elastic constants (38.27). The IPS display voltage depends only on K_2 , and for VAN and TVAN devices, the threshold voltage is determined by K_3 . Different combinations of elastic constants determine the transmission/voltage curves, which are important in the multiplexing of complex displays. For example, decreasing the ratio K_3/K_1 increases the steepness of the curve for TN displays, and so increases the number of lines that may be addressed. On the other hand, for the STN display, if the ratio K_3/K_1 is decreased, the number of lines that may be addressed also decreases. Identification of the important elastic

constants necessary to optimise the operation of these displays is relatively straightforward; however, manipulation of the components of displays mixtures to give the best results is much more difficult.

Ferroelectric and Flexoelectric Properties

The electro-optic properties considered so far result from interaction of an electric field with the anisotropic permittivity of a material. This might be termed a quadratic response since the dielectric term in the free energy (38.2) is quadratic in the electric field, and as a consequence the electro-optic response does not depend on the sign of the electric field. For materials having a centre of symmetry, such as achiral nematic and smectic liquid crystals, this response is the only one possible. However, if the centro-symmetry of the liquid crystal is broken in some way, then a linear electric polarisation becomes possible, which results in a linear response to an applied electric field. One example of this, in the context of displays, is the chiral smectic C phase, which in the surface-stabilised state exhibits ferroelectricity i.e. a spontaneous electric polarisation. The origin of the symmetry breaking in this case is the chirality of the material, and the polarisation is directed along an axis perpendicular to the tilt plane of the smectic C. Another way in which the symmetry can be broken is through elastic strain. This effect was first described by Meyer [38.24], and it can be represented as a polarisation resulting from a splay or bend deformation (Sect. 38.2.4). Since at a molecular level, strain is related to molecular interactions, the flexoelectric response depends on a coupling of intermolecular forces and the molecular charge distribution. Two molecular mechanisms have been identified which contribute to the strain-induced polarisation. If the molecules have a net dipole moment, then the longitudinal component can couple with the molecular shape to give a splay polarisation along the director axis, while the transverse component couples with the shape to give a bend polarisation perpendicular to the director axis. Even in the absence of a net dipole, a quadrupolar charge distribution in a molecule can result in strain-induced polarisation [38.26]. Both contribute to the splay and bend flexoelectric coefficients, but only the dipolar part persists in the sum $e_1 + e_3$. Thus it is common to quote flexoelectric coefficients as a sum and difference rather than as separate coefficients.

The measurement of flexoelectric coefficients has been a challenge to experimentalists, and there is a wide range of values in the literature for standard materials (Sect. 38.2.4). It is, therefore, premature to

draw any conclusions about structure/property relationships for flexoelectricity from the limited experimental data available. There have been attempts [38.73, 74] to model flexoelectricity for collections of Gay–Berne particles simulating wedge-shaped molecules. Application of the surface interaction model to flexoelectric behaviour [38.75] has allowed the calculation of flexoelectric coefficients for a number of molecules; these calculations include the quadrupolar contribution. The importance of molecular shape is clearly demonstrated, and in particular changes of shape, either through conformational changes or *cis–trans* isomerisation, have large effects on the magnitude and sign of the flexoelectric coefficients.

Flexoelectric effects contribute to the electro-optic response of nematic displays, especially those with hybrid alignment, i.e. planar on one electrode and homeotropic on the other electrode, but they are not usually considered in the optimisation of mixture properties. However, flexoelectric properties are of direct importance to the operation of displays based on the switching of the direction of the optic axis in chiral nematics: the so-called deformed helix mode [38.25].

Viscous Properties

As explained in Sect. 38.2.5, the flow properties of liquid crystals are complicated. Since the materials are anisotropic, the viscosities in different directions are different, furthermore because of the torsional elasticity, viscous stress can couple with the director orientation to produce complex flow patterns. Thus there are five viscosity coefficients necessary for nematics, in addition to the elastic constants, and as many as 20 viscosities for smectic C liquid crystals [38.76]. To relate all or indeed any of these to molecular structure is a formidable challenge. However, for most liquid-crystal displays, the only viscosity of interest is that which relates to the reorientation of the director: the so-called rotational viscosity. This depends on the temperature and order parameter, and on the forces experienced by the rotating director. The rotational viscosities for all liquid-crystalline materials can be represented by one or other of the following parameterised relations

$$\begin{aligned}\gamma_1 &= aS^x \exp\left(\frac{A(T)}{k_B T}\right) \\ \text{or} \\ \gamma_1 &= bS^y \exp\left(\frac{B}{T - T_0}\right),\end{aligned}\quad (38.35)$$

Table 38.4 Fluorinated mesogens of positive and negative dielectric anisotropy used in liquid-crystal mixtures for modern displays. All measurements listed have been obtained by extrapolation from measurements on nematic solutions

Compound Transition temperatures (°C) [ref.]	$\Delta\epsilon$	Δn	Rotational Viscosity (γ_1 /m Pa s)
 T_{CrN} 49 °C T_{NI} (13) °C [38.59]	-6.2	0.1	$\gamma_1 = 110$
 T_{CrN} 80 °C T_{NI} 173 °C [38.59]	-5.9	0.156	$\gamma_1 = 233$
 T_{CrN} 30 °C T_{NI} 58 °C [38.54]	11.3	0.134	$\gamma_1 = 191$ [38.77]
 T_{CrSmB} 43 °C T_{SmBN} 128 °C T_{NI} 147 °C [38.54]	9	0.14	$\gamma_1 = 180$ [38.77]
 T_{CrN} 67 °C T_{NI} 145 °C [38.59]	-2.7	0.095	$\gamma_1 = 218$
 T_{CrN} 88 °C T_{NI} 89 °C [38.59]	-4.3	0.2	$\gamma_1 = 210$
 T_{CrN} 56 °C T_{NI} 117 °C [38.59]	11.1	0.067	$\gamma_1 = 175$

where a , b , A and B are material parameters, and S is the order parameter raised to a power of x or y between 0 and 2. The first of these expressions emphasises the diffusional nature of rotational relaxation in a liquid

crystal, that is molecules rotating in an external potential. The second expression taken from polymer physics describes rotation in terms of free volume, where T_0 is the temperature at which the free volume becomes zero, and the rotational viscosity infinite i. e. a glass transition.

At the simplest level, the rotational viscosity depends on the molecular shape and size. As the length of the molecule increases, from two rings to three rings etc. γ_1 increases, similarly it increases with the length of the alkyl chain. Varying the nature of the rings in the mesogenic core can have a dramatic effect on the rotational viscosity, which correlates well with free volume, as shown in Fig. 38.25, and the glass temperature of the material.

There is a correlation between the increasing dielectric anisotropy and increasing rotational viscosity which can be attributed to local dipolar intermolecular forces which impede end-over-end rotation of molecules. Thus mesogens having cyano-groups in a terminal position tend to have relatively high viscosities. Other dipolar groups such as fluorine do not have such a deleterious effect on viscosity as the cyano-group, and so fluorine-containing mesogens are increasingly preferred in the formulation of display mixtures.

As with the determination of other properties, the rotational viscosities of mesogenic components are often determined by extrapolation from measurements on mixtures doped with the component under investigation. Such a method only provides approximate values to compare different components, but in the formulation of mixtures for display applications it is only the rotational viscosity of the final mixture that is important. Rotational viscosities of some mesogens of interest for display mixtures are given in the tables accompanying this section. In many instances, the viscosities, as with other properties, have been determined by extrapolation from measurements on mixtures. The measurement of rotational viscosities is experimentally difficult, and some authors prefer to quote the results of bulk-viscosity measurements in terms of a kinematic viscosity. In fact there is a good correlation between kinematic viscosity and rotational viscosity, and where

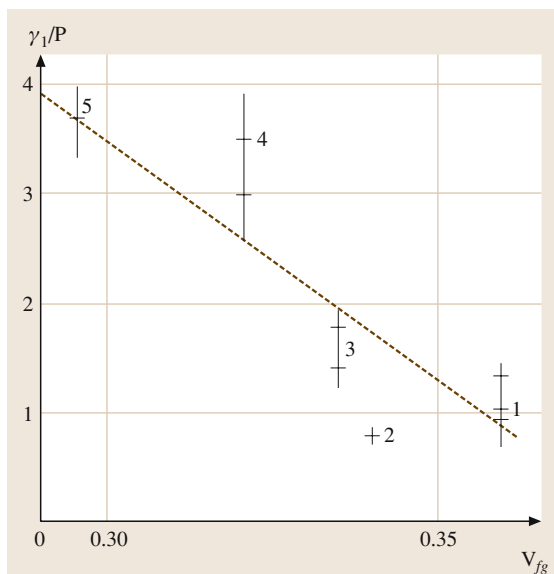


Fig. 38.25 Relationship between the rotational viscosity coefficient γ_1 (Poise = 10^{-1} Pa s) and the geometric free volume (V_{fg}) at 25 °C for bicyclic polar mesogens. Compound numbers represent alkyl series as follows: 1-cyanophenylalkylcyclohexanes; 2-alkylcyanobiphenyls; 3-cyanophenylalkylpyridines; 4-cyanophenylalkylbicyclooctanes; 5-alkoxycyanobiphenyls (After 38.52)

possible both values have been included in the tables. From the various tables, the effect of increasing the molecular length on the viscosity is clearly seen, as is the effect of replacing an F atom with a CN group. Lateral substitution, which produces materials of negative dielectric anisotropy, tends to increase the rotational viscosity. Despite this the fastest nematic displays now use vertical alignment and materials of negative dielectric anisotropy. The operating characteristics of display mixtures depend on the physical properties of individual components in a very complex manner, and optimisation of mixture properties has to be carried out in a concerted way.

References

- 38.1 P. J. Collings, M. Hird: *Introduction to Liquid Crystals* (Taylor & Francis, London 1997)
- 38.2 M. A. Bates, G. R. Luckhurst: Computer simulation studies of anisotropic systems. XXX. The phase behaviour and structure of a Gay-Berne mesogen, *J. Chem. Phys.* **110**, 7087–7108 (1999)
- 38.3 M. Hird: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 1.1
- 38.4 P. Pasini, C. Zannoni: *Advances in the Simulation of Liquid Crystals* (Kluwer, Dordrecht 1998)

- 38.5 J. W. Emsley, G. R. Luckhurst, G. H. Shilstone, I. Sage: The preparation and properties of the α,ω -bis(4,4'-cyanobiphenyloxy)alkanes: nematogenic molecules with a flexible core, *Mol. Cryst. Liq. Cryst. Lett.* **110**, 223–233 (1984)
- 38.6 G. W. Gray: *The Molecular Physics of Liquid Crystals*, ed. by G. R. Luckhurst, G. W. Gray (Academic, New York 1979) Chap. 12
- 38.7 T. J. Sluckin, D. A. Dunmur, H. Stegemeyer: *Crystals that Flow* (Taylor & Francis, New York 2004)
- 38.8 R. B. Meyer, L. Liebert, L. Strzelecki, P. J. Keller: Ferroelectric liquid crystals, *J. Phys. (Paris)* **L36**, 69–71 (1975)
- 38.9 S. R. Renn, T. C. Lubensky: Abrikosov dislocation lattice in a model of the cholesteric-to-smectic A transition, *Phys. Rev. A* **38**, 2132–2147 (1988)
- 38.10 J. W. Goodby, M. A. Waugh, S. M. Stein, E. Chin, R. Pindak, J. S. Patel: Characterization of a new helical smectic liquid crystal, *Nature* **337**, 449–451 (1989)
- 38.11 S. Chandrasekhar, B. K. Sadashiva, K. A. Suresh: Liquid crystals of disc-like molecules, *Pramana* **9**, 471–480 (1977)
- 38.12 K. Praefcke: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 1.2
- 38.13 G. G. Nair, D. S. S. Rao, K. S. Prasad, S. Chandrasekhar, S. Kumar: Electrooptic and viewing angle characteristics of a display device employing a discotic nematic liquid crystal, *Mol. Cryst. Liq. Cryst.* **397**, 545–552 (2003)
- 38.14 N. Boden, R. Bissel, J. Clements, B. Movaghar: Discotic liquid crystals: self-organising molecular wires, *Liq. Cryst. Today* **6(1)**, 1–4 (1996)
- 38.15 V. Tsvetkov: Über die Molekulanordnung in der anisotrop-flüssigen Phase, *Acta Physicochim* **15**, 132–147 (1942)
- 38.16 G. R. Luckhurst: *Dynamics and Defects in Liquid Crystals*, ed. by P. E. Cladis, P. Palffy-Muhoray (Gordon & Breach, Philadelphia 1998) p. 141
- 38.17 P. G. de Gennes: *The Physics of Liquid Crystals* (Oxford Univ. Press, Oxford 1974) p. 96
- 38.18 J. Cognard: Alignment of nematic liquid crystals and their mixtures, *Mol. Cryst. Liq. Cryst.* **1**, 1–77 (1982), (Suppl.)
- 38.19 D. W. Berreman: Solid surface shape and the alignment of an adjacent nematic liquid crystal, *Phys. Rev. Lett.* **28**, 1683–1686 (1972)
- 38.20 A. Rapini, M. Papoular: Distortion d'une lamelle nématique sous champ magnétique. Conditions d'anchorage aux parois, *J. Phys. Colloq. (France)* **30**, C-4–54–56 (1969)
- 38.21 A. Sugimura: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 10.2
- 38.22 F. C. Frank: On the theory of liquid crystals, *Trans. Faraday Soc.* **25**, 19–28 (1958)
- 38.23 V. Fréedericksz, V. Zolina: Forces causing the orientation of an anisotropic liquid, *Trans. Faraday Soc.* **29**, 919–930 (1933)
- 38.24 R. B. Meyer: Piezoelectric effects in liquid crystals, *Phys. Rev. Lett.* **22**, 918–921 (1969)
- 38.25 P. Rudquist, S. T. Lagerwall: On the flexoelectric effect in nematics, *Liq. Cryst.* **23**, 503–510 (1997)
- 38.26 A. E. Petrov: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 5.5
- 38.27 J. Prost, J. P. Marcerou: On the microscopic interpretation of flexoelectricity, *J. Phys. (Paris)* **38**, 315–324 (1977)
- 38.28 J. K. Mosicki: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 8.2
- 38.29 W. Helfrich: Molecular theory of flow alignment of nematic liquid crystals, *J. Chem. Phys.* **50**, 100–106 (1969)
- 38.30 C. Gähwiller: The viscosity coefficients of a room-temperature liquid crystal (MBBA), *Mol. Cryst. Liq. Cryst.* **20**, 301–318 (1973)
- 38.31 H. Knepe, F. Schneider, N. K. Sharma: Rotational viscosity γ_1 of nematic liquid crystals, *J. Chem. Phys.* **77**, 3203–3208 (1982)
- 38.32 F. M. Leslie: *The Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 8.1
- 38.33 O. Parodi: Stress tensor for a nematic liquid crystal, *J. Phys. (Paris)* **31**, 581–584 (1970)
- 38.34 Y. Björnsthål: Untersuchungen über die anisotropen Flüssigkeiten, *Ann. der Phys.* **56**, 161–207 (1918)
- 38.35 B. Levin, N. Levin: Improvements in or relating to light valves, *British Patent* 441,274 (1936)
- 38.36 P. Yeh, C. Gu: *Optics of Liquid Crystal Displays* (Wiley, New York 1999)
- 38.37 P. M. Alt, P. Pleshko: Scanning limitations of liquid crystal displays, *IEEE Trans. Electron. Dev.* **21**, 146–155 (1974)
- 38.38 E. Jakeman, E. P. Raynes: Electro-optic response times in liquid crystals, *Phys. Lett.* **39A**, 69–70 (1972)
- 38.39 K. Tarumi, U. Finkenzeller, B. Schuler: Dynamic behaviour of twisted nematic cells, *Jpn. J. Appl. Phys.* **31**, 2829–2836 (1992)
- 38.40 M. Schadt, W. Helfrich: Voltage-dependent optical activity of a twisted nematic liquid crystal, *Appl. Phys. Lett.* **18**, 127–128 (1971)
- 38.41 Ch. Mauguin: Sur les cristaux liquides de Lehmann, *Bull. Soc. Fran. Mineral.* **34**, 71–117 (1911)
- 38.42 C. H. Gooch, H. A. Tarry: The optical properties of twisted nematic liquid crystal structures with twist angles $\leq 90^\circ$, *J. Phys. D: Appl. Phys.* **8**, 1575–1584 (1975)
- 38.43 T. J. Scheffer, J. Nehring: A new highly multiplexed liquid crystal display, *Appl. Phys. Lett.* **45**, 1021–1023 (1984)

- 38.44 C. M. Waters, E. P. Raynes, V. Brimmell: Design of highly multiplexed liquid crystal dye displays, *Mol. Cryst. Liq. Cryst.* **123**, 303–319 (1985)
- 38.45 A. J. Snell, K. D. Mackenzie, W. E. Spear, P. G. LeComber, A. J. Hughes: Application of amorphous silicon field effect transistors in addressable liquid crystal display panels, *Appl. Phys.* **24**, 357–362 (1981)
- 38.46 G. Baur, R. Kiefer, H. Klausmann, F. Windscheid: In-plane switching: a novel electro-optic effect, *Liq. Cryst. Today* **5**(3), 13–14 (1995)
- 38.47 M. Oh-e, K. Kondo: Electro-optical characteristics and switching behaviour of the in-plane switching mode, *Appl. Phys. Lett.* **67**, 3895–3897 (1966)
- 38.48 N. A. Clark, S. T. Lagerwall: Submicrosecond bistable electro-optic switching in liquid crystals, *Appl. Phys. Lett.* **36**, 899–901 (1980)
- 38.49 G. P. Bryan-Brown, C. V. Brown, I. C. Sage, V. C. Hui: Voltage-dependent anchoring of a liquid crystal on a grating surface, *Nature* **392**, 365–367 (1998)
- 38.50 C. V. Brown, L. Parry-Jones, S. J. Elston, S. J. Wilkins: Comparison of theoretical and experimental switching curves for a zenithally bistable nematic liquid crystal device, *Mol. Cryst. Liq. Cryst.* **410**, 417–425 (2004)
- 38.51 J. S. Patel, R. B. Meyer: Flexoelectric electro-optics of a cholesteric liquid crystal, *Phys. Rev. Lett.* **58**, 1538–1540 (1987)
- 38.52 V. V. Belyaev: *Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 8.4
- 38.53 G. W. Gray, S. M. Kelly: Liquid crystals for twisted nematic displays, *J. Mater. Chem.* **9**, 2037–2050 (1999)
- 38.54 V. F. Petrov: Liquid crystals for AMLCD and TFT-PDLC applications, *Liq. Cryst.* **19**, 729–741 (1995)
- 38.55 H. Takatsu, K. Takeuchi, H. Sato: *Mol. Cryst. Liq. Cryst.* **100**, 345–355 (1983)
- 38.56 D. Coates: In: *Liquid Crystals, Applications and Uses*, Vol. 1, ed. by B. Bahadur (World Scientific, Singapore 1990) p. 91
- 38.57 L. Pohl, U. Finkelzeller: In: *Liquid Crystals, Applications and Uses*, Vol. 1, ed. by B. Bahadur (World Scientific, Singapore 1990) p. 1139
- 38.58 K. Tarumi, M. Bremer, T. Geelhaar: Recent liquid crystal material development for active matrix displays, *Ann. Rev. Mater. Sci.* **27**, 423–441 (1997)
- 38.59 D. Pauluth, K. Tarumi: Advanced liquid crystals for television, *J. Mater. Chem.* **14**, 1219–1227 (2004)
- 38.60 D. A. Dunmur, A. Fukuda, G. R. Luckhurst (Eds.): *Physical Properties of Liquid Crystals: Nematics* (INSPEC, London 2001)
- 38.61 D. S. Hulme, E. P. Raynes, K. J. Harrison: Eutectic mixtures of nematic 4'-substituted 4-cyanobiphenyls, *J. Chem. Soc. Chem. Comm.*, 98–99 (1974)
- 38.62 D. Demus, Ch. Fietkau, R. Schubert, H. Kehlen: Calculation and experimental verification of eutectic systems with nematic phases, *Mol. Cryst. Liq. Cryst.* **25**, 215–232 (1974)
- 38.63 R. L. Humphries, P. G. James, G. R. Luckhurst: A molecular field treatment of liquid crystalline mixtures, *Symp. Faraday Trans.* **5**, 107–118 (1971)
- 38.64 G. H. Heilmeyer, L. A. Zannoni, L. A. Barton: Dynamic scattering: a new electro-optic effect in certain classes of nematic liquid crystals, *Proc. IEEE* **56**, 1162–1171 (1968)
- 38.65 D. A. Dunmur, K. Toriyama: In: *Physical Properties of Liquid Crystals*, ed. by D. Demus, J. Goodby, G. W. Gray, H.-W. Spiess, V. Vill (Wiley-VCH, Weinheim 1999) p. 129
- 38.66 D. A. Dunmur: *Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 7.1
- 38.67 D. A. Dunmur, M. R. Manterfield, W. H. Miller, J. K. Dunleavy: The dielectric and optical properties of the homologous series of cyano-alkylbiphenyl liquid crystals, *Mol. Cryst. Liq. Cryst.* **45**, 127–144 (1978)
- 38.68 K. Toriyama, D. A. Dunmur, S. E. Hunt: Transverse dipole association and negative dielectric anisotropy of nematic liquid crystals, *Liq. Cryst.* **5**, 1001–1009 (1989)
- 38.69 D. A. Dunmur, D. A. Hitchen, X.-J. Hong: The physical and molecular properties of some nematic fluorobiphenylalkanes, *Mol. Cryst. Liq. Cryst.* **140**, 303–318 (1986)
- 38.70 S. Naemura: *Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 11.2
- 38.71 H. Gruler: The elastic constants of a nematic liquid crystal, *Z. Naturforsch.* **30a**, 230–234 (1975)
- 38.72 R. G. Priest: Theory of the Frank elastic constants of nematic liquid crystals, *Phys. Rev. A* **7**, 720–729 (1973)
- 38.73 J. Stelzer, R. Beradi, C. Zannoni: Flexoelectric effects in liquid crystals formed by pear-shaped molecules. A computer simulation study, *Chem. Phys. Lett.* **299**, 9–16 (1999)
- 38.74 J. L. Billeter, R. A. Pelcovits: Molecular shape and flexoelectricity, *Liq. Cryst.* **27**, 1151–1160 (2000)
- 38.75 A. Ferrarini: Shape model for the molecular interpretation of the flexoelectric effect, *Phys. Rev. E* **64**, 021710–11 (2001)
- 38.76 I. W. Stewart: *The Static and Dynamic Continuum Theory of Liquid Crystals* (Taylor & Francis, London 2004)
- 38.77 K. Tarumi, M. Heckmeier: *Physical Properties of Liquid Crystals: Nematics*, ed. by D. A. Dunmur, A. Fukuda, G. R. Luckhurst (INSPEC, London 2001) Chap. 11.4

Organic Photo

39. Organic Photoconductors

This Chapter surveys organic photoreceptor devices used in electrophotography. Included in the discussion are the materials (polymers, pigments, charge-transport molecules, etc.), device architecture, fabrication methods, and device electrical characteristics that are critical to the successful functioning of an electrophotographic device (printer).

The Chapter is organized as follows. A brief discussion of the history of xerography and the contributions of Chester Carlson is followed by operational considerations and critical materials properties. The latter includes dark conductivity, photodischarge-charge transport, and photogeneration. Organic photoreceptor characterizations of dark decay, photosensitivity, and electrical-only cycling are discussed in detail. This is followed by discussions of photoreceptor architecture, coating technologies, substrate, conductive layer, and coated layers which carry out specific functions such as smoothing, charge blocking, charge transport, backing, and surface protection.

39.1	Chester Carlson and Xerography	954
39.2	Operational Considerations and Critical Materials Properties	956
39.2.1	Dark Conductivity	956
39.2.2	Photodischarge-Charge Transport	957
39.2.3	Photogeneration	963
39.3	OPC Characterization	965
39.3.1	Dark Decay	965
39.3.2	Photosensitivity	965
39.3.3	Electrical-Only Cycling	966
39.4	OPC Architecture and Composition	967
39.4.1	OPC Architecture	967
39.4.2	Coating Technologies	968
39.4.3	Substrate and Conductive Layer ..	969
39.4.4	Smoothing Layer and Charge-Blocking Layer	969
39.4.5	Charge-Generation Layer (CGL)	970
39.4.6	Charge-Transport Layer (CTL)	974
39.4.7	Backing Layer	975
39.4.8	Overcoat Layer	975
39.5	Photoreceptor Fabrication	976
39.6	Summary	977
	References	978

Organic photoconductors, devices fabricated from organic photoconductors, and the applications of these devices, are the topics which will be covered in this chapter. The term *organic* is used in the chemical sense to encompass materials with carbon as a major constituent. This includes molecular as well as polymeric materials that are not naturally occurring but are purposefully designed and synthesized in the laboratory for specific physical, chemical, dark-electrical, and photoelectrical characteristics. Organic photoconductors are single, or more commonly, a formulated blend of materials which have photoconductive characteristics. Put simply, these are materials in which the electrical conductivity increases on exposure to light [39.1, 2]. Organic photoreceptors are thin-film devices made from organic photoconductive materials with physical, chemical, dark-electrical, and photoelectrical characteristics designed for optimum performance in

specific applications. The most important application is in electrophotography, where organic photoreceptors are utilized in machines for digital printing and copying. In this chapter we will discuss details of organic photoreceptor architecture and their composition and characteristics that have enabled their utilization in electrophotographic printing. In this space it will only be possible to touch briefly on many important aspects of organic photoreceptors but references will be provided for those wishing to delve deeper [39.3–6].

Organic photoconductors are materials and organic photoreceptors are devices made from organic photoconductors. The common acronym for an organic photoreceptor is OPC, which of course stands for organic photoconductor, and simply indicates that acronyms are often inscrutable. The application for which organic photoreceptors have received the most attention and development effort is as the photosensitive

image-creating element in electrophotographic-based printers and copiers. In these machines the photoreceptor is a highly uniform, defect-free, multilayer thin film ($\approx 20\text{--}50\text{ }\mu\text{m}$ photoconductor thickness) coated on a metal drum or a flexible belt, with an active area

from 160 to over $12\,000\text{ cm}^2$. Organic photoreceptors for electrophotography were first introduced in the 1960s and, after years of development for improved physical and electrical characteristics, they form the basis of a multi-billion-dollar industry today.

39.1 Chester Carlson and Xerography

Electrophotography has an interesting history [39.7–10] and we will briefly tell this story with emphasis on

the enabling aspects of organic photoreceptors in this technology. The story begins with Chester F. Carlson, a patent lawyer who saw the need for a method of reproducing documents. In his kitchen laboratory in Queens, New York, he demonstrated that a charged photoconductor could be used to produce an electrostatic latent image which could subsequently be visualized by contact with oppositely charged particles. The first electrophotographic image was “10.-22.-38 Astoria”. The patent application was filed on 4th April 1939 and on 6th October 1942, US patent 2 297 691 was issued [39.11]. Figure 39.1 shows the title page of that patent. In concept, electrophotography is a relatively simple process with six (or seven) steps but it took over 20 years to bring it to fruition and another decade to make it truly practical. The essential steps are all illustrated in the US patent cover page: charging, exposure, developing, transfer and fixing. Not pictured are the final steps, blanket exposure and cleaning, to prepare the photoreceptor for the next electrophotographic cycle.

Carlson had great difficulty finding commercial backing for his invention. He was unable to convince the management at a long list of influential and affluent companies that making copies of documents was an attractive business proposition. In the end the Battelle Development Corporation took on the project and made significant advances towards making it more practical. In Carlson’s invention a coating of photoconductive material such as sulfur or anthracene is prepared on a conductive, grounded, metal substrate. The photoconductor is tribocharged by contact with an appropriate material (silk rabbit’s fur, etc.). The photoconductor is then image-wise exposed with light of a wavelength that causes the photoconductive response. The areas struck by light become conductive and the surface charge is neutralized. Thus, there is now a difference in surface charge density, and hence surface potential, between the exposed and unexposed areas. This difference in surface potential, the so-called electrostatic latent image, was visualized by contacting the surface with small marking particles, such as spores from the lycopodium club

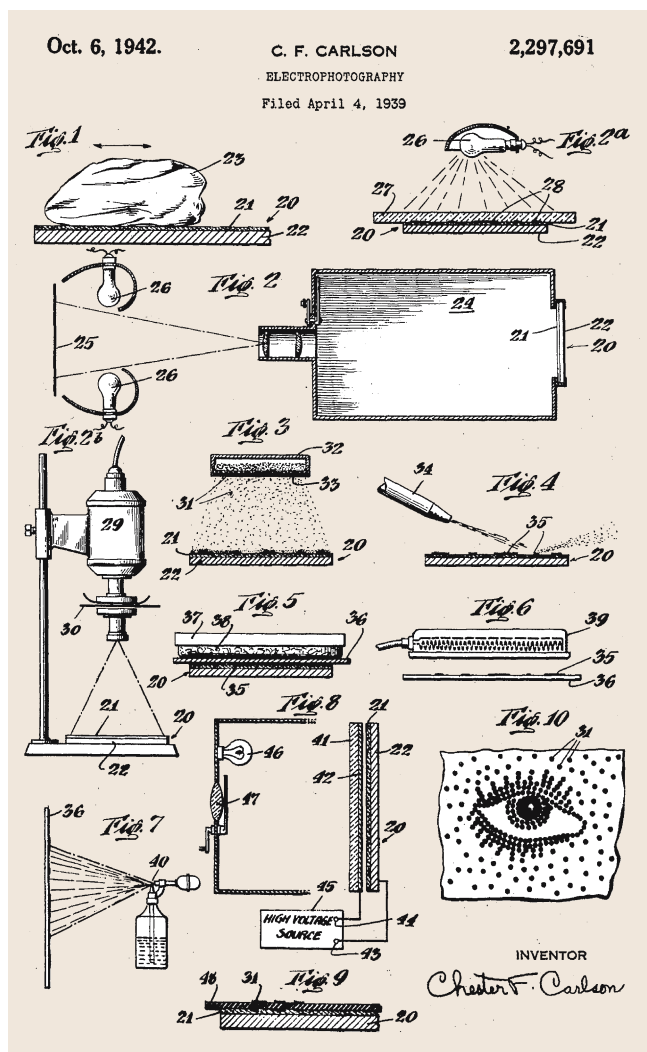


Fig. 39.1 Title page of the first U. S. Patent on electrophotography by C. F. Carlson

moss, with surface charge opposite to the original surface charge. The final copy is produced by transferring the *toned* image on the photoreceptor to the receiver (Carlson accomplished this with a heated sheet of waxed paper). Carlson later dubbed his process *xerography*, meaning dry writing, based on the use of dry powder as the marking particles. Electrophotography is a more general term encompassing, for example, the use of liquid toners in which the charged marking particles are suspended in an insulating fluid.

Two significant early advances were the use of amorphous selenium as the photoconductive material and the development of a practical method (the corotron) for the generation of an air corona for surface charging the photoconductive material. The corotron is simply a fine wire held at high potential (above the air ionization threshold of ≈ 7 kV) and shielded on the back with a grounded metal plate. Xerography was eventually purchased by the Haloid Corporation of Rochester, New York, (later to become the Xerox Corporation) with Carlson as a consultant. The first xerographic product was the XeroX Copier model A introduced in 1949. This initial product had totally manual operation which required several minutes to produce a copy. The photoreceptor was a selenium plate that measured 9×14 inches. From these modest beginnings, advances in engineering the various processes and materials progressed in an evolutionary manner leading to the introduction of the automated copy machine, the Xerox 914 in 1959. This utilized a selenium-coated drum photoreceptor. The evolution continues today and new product introductions have rapidly shifted from copiers (reflection of light from a document to expose the photoreceptor) to digital printers (pixel-wise exposure of the photoreceptor based on data from a previously compiled image file) using drum- or flexible-belt-based organic photoreceptors. Early examples of copiers utilizing organic photoreceptors are the IBM copier 1 (1970) and the Kodak Ektaprint 100 (1975). Initially the photoreceptors were *single-layer* in that the photoconductive materials were all contained in a single layer. For high-speed high-quality applications more complex photoreceptors were introduced. These had multilayer structures with each layer being optimized for a specific function. The details of photoreceptor architecture and composition will be discussed later.

There are some subtle differences in the electrophotographic process between copiers and digital printers. Since the electrostatic image in a copier is produced by reflection of the exposing light from a document (black text for example), areas of the document which are dark

(text) reflect no light and the corresponding area of the photoreceptor retains its surface charge. To produce a final print in which the dark areas correspond to dark areas in the original document one must *develop* the image with a marking particle (toner) with charge that is opposite to that on the photoreceptor surface. This process is called *charged area development*. In a digital printer the exposure is controlled by an image file and the image is composed of a microscopic halftone pattern of *pixels*. Pixel size is determined by the desired *resolution* such that 600 dpi corresponds to a pixel size of about $40 \mu\text{m}$. In a digital process it is possible to expose either those areas which will be developed or undeveloped in the final print. The approach sometimes chosen is to expose only those areas which will be toned since most documents are text-based and text occupies only a small fraction of the total document area. This approach decreases the on-time of the exposing system and lengthens its lifetime. Thus, in this case those areas to be developed have been exposed and have decreased surface potential. This process, *discharged area development*, is carried out by using toner with a charge polarity that is the same as that of the photoreceptor surface. Factors in OPC design relating to their optimization for digital imaging have been investigated but in practice the OPCs used in today's copiers and printers differ little [39.12–14]. Discussion of the chemistry, physics, and engineering involved in toners, developers, and development systems is beyond the scope of this chapter [39.15, 16].

In addition to the photoreceptor some of the major subsystems of a modern digital printer are:

- Exposure: scanning laser or light-emitting diode (LED) arrays.
- Development: magnetic brush or other technology with black and/or colored marking particles plus magnetic carrier.
- Transfer: heat, pressure, etc. to remove toner from the photoreceptor and place it on the receiver. In some systems the toner is transferred first to an intermediate drum or web and then to the final receiver.
- Erase: blanket exposure to return the photoreceptor to a uniform surface-charge state.
- Clean: blade or brush to remove untransferred toner from the photoreceptor surface.
- Fixing: melting the toner particles onto the receiver.
- Computers and software: image capture, rendering, and storage, process control, receiver handling, etc.

- Process control: software and hardware (sensors, voltmeters, etc.) to maintain image quality.
- Sensors and controls: for receiver handling.

A full listing would be much longer and the modern electrophotographic printer is a highly complex system where the hardware, software, and materials have been successfully co-optimized to meet the product aims [39.17–19].

Since the photoreceptor surface is either in physical contact with (development, transfer, cleaning), or exposed to, effluents (fusing, corona charging) from the various subsystems its physical and electrical characteristics must be stable to these interactions throughout its life. Much of the development in OPCs has been to extend the photoreceptor process lifetime. Today the most durable OPCs might be replaced after 100 000 or more imaging cycles.

39.2 Operational Considerations and Critical Materials Properties

The steps involved in the electrophotographic process were detailed previously [39.15, 17]. In this section we discuss some of the underlying physics in the formation of the electrostatic latent image. From the point of view of device physics the OPC is a large-area transducer configured as a belt or drum which converts optical information into a latent charge (primary) image. The photoreceptor material is a high-dielectric-strength insulator that is converted to the electrically conducting state with the application of a field and illumination. Phenomenologically, a high field is applied across the thin-film photoconductor (corona charging) and the film is exposed to radiation absorbed by the material. The photon energy is converted into charge carriers at or near the site of absorption and these drift under the influence of the field. Since the charged photoreceptor is an open-circuit device the surface potential decreases as the charges drift through the material and the photoreceptor is discharged.

In this section we focus on some of the key physics and materials issues governing the field-biased motion of electronic charge through the photoreceptor during light-induced xerographic discharge – the process which leads to latent image formation. Work done in optimizing the design and characterization of photoreceptors in order to service an evolving technology operated in a push–pull relationship with the growth in scientific understanding of photoinduced charge generation, injection, transport and trapping first in amorphous semiconductors and then, in what will concern us here, disordered molecular media.

Because the photoreceptor assembly is a sensitized large-area device, the materials which simultaneously optimize all the required properties have always been glassy coatings. (Polycrystalline media are largely unsuitable for a variety of reasons.) The inorganic materials used especially in light-lens copiers were initially amor-

phous chalcogenides [39.20, 21] and to a much lesser extent hydrogenated amorphous Si [39.22, 23]. Much of the early work was therefore focused on the physics and chemistry of amorphous semiconductors. In this case photogeneration and the subsequent transport of charge occur in precisely the same medium, and design latitude is clearly restricted by the particular combination of optical and transport characteristics of that single layer. In addition these inorganic films are relatively brittle and therefore unsuitable for applications which require a belt architecture, i.e., one in which the photoreceptor is required to bend around small rollers. On the other hand, polymer-based OPCs are inherently flexible and multi-layer architectures are readily fabricated such that each layer can be optimized for a particular function such as photogeneration and charge transport [39.24]. In fact, all OPCs for high-end applications have been developed according to this principle with the photoconductor divided into a thin light-absorbing charge-generation layer (CGL) adjacent to the electrode and a thicker charge-transport layer (CTL) which transports holes. With this photoreceptor configuration the surface is negatively charged prior to exposure. The details of OPC architecture and the materials chosen for specific layer functions will be discussed later.

39.2.1 Dark Conductivity

The electrophotographic imaging process begins by applying a surface charge to the photoreceptor. A parallel-plate-capacitor model is appropriate and the surface potential and surface charge density are related through the capacitance per unit area as, $Q/A = (C/A)V = (\epsilon\epsilon_0/L)V$, where Q/A is the surface charge density (C/cm^2), C/A is the capacitance per unit area (F/cm^2), V is the surface potential (V), ϵ is the dielectric constant, ϵ_0 is the permittivity

of free space, and L is the photoreceptor thickness. For a typical organic photoreceptor ($L = 25\text{ }\mu\text{m}$, $\epsilon = 3$, and $V = -500\text{ V}$) C/A is $3.2 \times 10^{-10}\text{ F/cm}^2$, corresponding to 10^{12} charges/ cm^2 . Assuming the surface is composed of molecules with an area of $100\text{ }\text{\AA}^2$ ($10^{-14}\text{ cm}^2/\text{molecule}$) only $\approx 1\%$ of the surface molecules are associated with the surface charge. The applied field (V/L) is $2 \times 10^5\text{ V/cm}$.

Electrophotographic imaging is enabled by a difference in surface potential between exposed and unexposed areas when the imaged area enters the development subsystem. Generally a potential difference of at least 300 V is desired. The materials comprising the bulk of the OPC must be highly insulating. A resistivity of $10^{13}\text{ }\Omega\text{ cm}$ at a field of $2 \times 10^5\text{ V/cm}$ will give rise to a dark decay rate of nearly 200 V/s . Since development typically occurs at a fraction of a second to one second after charging, the photoconductor resistivity needs to be $> 10^{13}\text{ }\Omega\text{ cm}$. Organic polymers such as bisphenol-A polycarbonate meet this requirement and are the major component in OPCs.

There are other sources of dark conductivity such as electrode and/or surface charge injection, bulk charge generation, and/or charge detrapping [39.25]. Electrode injection is prevented with the interposition of a charge-injection blocking layer between the electrode and the photoconductive material. Surface injection is prevented by the chemical composition of the surface layer (which is typically a hole-transporting CTL). Detrapping can be a significant source of dark decay in OPCs and must be controlled by a balance between the process and the materials (the charge-generation material in particular).

39.2.2 Photodischarge–Charge Transport

The photoreceptor must retain charge in the dark and also be photosensitive enough to discharge exposed areas to half their initial charge potential when irradiated with (nominally for a mid-volume laser printer engine) $4\text{--}10\text{ erg/cm}^2$. The xerographic gain or quantum efficiency of supply describes the fractional number of surface charges neutralized per absorbed photon. For a dual-layer OPC it is a complicated convolution of the quantum efficiency of generation in the charge-generation layer, CGL, the efficiency of carrier injection from the CGL to the CTL and the transport parameters of the CTL. For this discussion we will for the moment ignore issues of charge generation and focus on transport. In dual-layer OPCs the majority of photodischarge occurs via charge transport through the CTL with the CGL playing a minor role. In the transport lexicon

[39.24,26] there are essentially two parameters that constitute the figures of merit to characterize charge motion through the polymeric CTL. These are (1) the drift mobility μ , the measure of *how fast* the carrier moves per unit applied field, and (2) the normalized carrier range $\mu\tau$ (τ is the free-carrier lifetime against deep trapping), which is *how far* the injected carrier moves per unit field before becoming immobilized in a deep trap. The time for a photoinjected carrier to traverse the CTL is called the transit time. The transit time t_{tr} , and mobility μ are related to the specimen thickness L according to $t_{\text{tr}} = L/\mu E$. The importance of mobility as a critical parameter in the electrophotographic process can be understood as follows: for an increase in the exposure intensity $d(F)$, the final decrease in surface potential $d(V)$ is proportional to the number of injected carriers and the distance they travel within the CTL. During xerographic discharge, a charge of CV_0 (C is the CTL capacitance and V_0 the initial voltage) traverses the bulk and induces time-dependent variation in the electric field behind the leading edge of the injected carrier front. Thus, as the fastest carriers transit the CTL, the electric field behind them is reduced, and the carriers behind the leading edge transit at a lower field, which in turn makes their velocities lower. Thus during xerographic discharge the transit times of individual photoinjected carriers become dispersed over a wide range, typically about an order of magnitude. For discharge to proceed to completion, even in the complete absence of deep trapping, enough time is required for the slowest carriers in the packet to exit the layer before the photoreceptor reaches the development zone – nominally $0.3\text{--}1.0\text{ s}$ after exposure in mid-volume printers. The latter must be allowed for in practice. Thus, carrier mobility in this particular illustration should exceed $10^{-6}\text{ cm}^2/\text{Vs}$. For example, consider that a dual-layer photoreceptor with a $25\text{-}\mu\text{m}$ CTL, in which there are no deep traps, is subjected to a light flash intense enough to ultimately induce complete discharge (CV_0 of absorbed photons). Consider further that in this CTL the mobility $\mu = 10^{-6}\text{ cm}^2/\text{Vs}$ at $E = 10^4\text{ V/cm}$ and that the device is initially charged to 1000 V . It can be calculated that under these conditions the device will have a residual voltage of 20 V after 0.3 s or 7 V after 1 s . Incomplete discharge, unless compensated for, might result in an inadequate toning potential and a toned density less than desired. Note further that, if the mobility is even lower, the results can become totally unacceptable. For example when the mobility is $\mu = 10^{-7}\text{ cm}^2/\text{Vs}$ under the conditions just described, the residual voltage a full second after exposure is 60 V , even in a completely trap-free CTL. In light of the fore-

going illustration for the trap-free case, which sets the mobility benchmark, the effect of traps in the polymeric CTL must be of paramount concern. More precisely, we are concerned with traps whose release time at ambient temperature discernibly exceeds the period of a complete electrophotographic cycle. In the present context we take the latter as the operational definition of a deep trap. With such traps present it is clearly the case that, after repeated charge-expose cycling, some quantity of image degrading charge would remain immobilized in the bulk for times now exceeding the period of a complete xerographic cycle. If ρ is the density of uniformly trapped space charge in a CTL of thickness L and relative dielectric constant ϵ then there is an associated residual potential V_R , where $V_R = e\rho L^2/2\epsilon\epsilon_0$. Here e is the electronic charge and ϵ_0 is the free-space permittivity. In a nonpolar dielectric medium ($\epsilon = 3$) with a layer thickness of 25 μm , as few as 10^{13} electronic charges trapped per cm^3 already give rise to a residual of 19 V. Unless process control utilizing electronic feedback correction can be employed, bulk-trapping-induced space-charge buildup during cycling can result in severe image degradation. Residual potential arising from bulk-trapped space charge is a critical electrophotographic process parameter. A residual potential can be related to the normalized carrier range, $\mu\tau$, in the weak-trapping limit from the physically plausible ansatz that the residual potential corresponds to that applied voltage for which the carrier range is about half the specimen thickness L . Thus, $\mu\tau$ is approximately $L^2/2V_R$. On this basis the tolerable trap density is defined by the requirement that the $\mu\tau$ product in practical devices should typically exceed $10^{-6} \text{ cm}^2/\text{V}$. Even highly purified polymer will typically contain many chemical impurities in the 1–10 ppm range, which the foregoing calculations show is vastly higher than permissible trap levels. However, chemical impurities, even when present at relatively high concentration, can be rendered trap-inactive by employing molecular design principles to guide the overall choice of active materials. Concepts derived from scientific understanding of photogeneration, injection across interfaces, and electronic transport in disordered organic materials, in combination with the unique compositional flexibility characteristic of the organic solid state, were together responsible for making multilayer OPCs the dominant practical receptor technology for electrophotography.

As discussed above, successful electrophotographic imaging requires that charge transport through the CTL occur with little trapping in the time scale of the electrophotographic cycle. Here we discuss how this

is accomplished in an *impure* organic-chemical-based CTL. The CTL is a glassy solid solution of a charge-transport-active moiety dispersed in a polymer binder. The charge-transport moiety can be molecular or a polymer component. The molecular solutions are called molecularly doped polymers (MDP). The transport-active component is typically 40–50% weight fraction of the CTL. Thus, *doping* in the present context is decidedly different from what doping refers to in conventional semiconductor physics [39.27]. In the latter case the dopant is typically introduced at very low concentration to control the relative proportion of electrons and holes in the bands of a semiconductor crystal leaving optical properties largely unaffected. In a semiconductor crystal at ambient temperature the dominant field driven transport mechanism is scattering perturbed motion of charge carriers in the bands. Under these circumstances, mobility decreases algebraically with increasing temperature, while carrier population in the bands is thermally activated. The convolution of these processes manifests in a thermally activated conductivity. On the other hand, in MDPs under equilibrium conditions the transport-active molecule is in the neutral state, and the glassy films are perfect insulators, that is, there are no free carriers

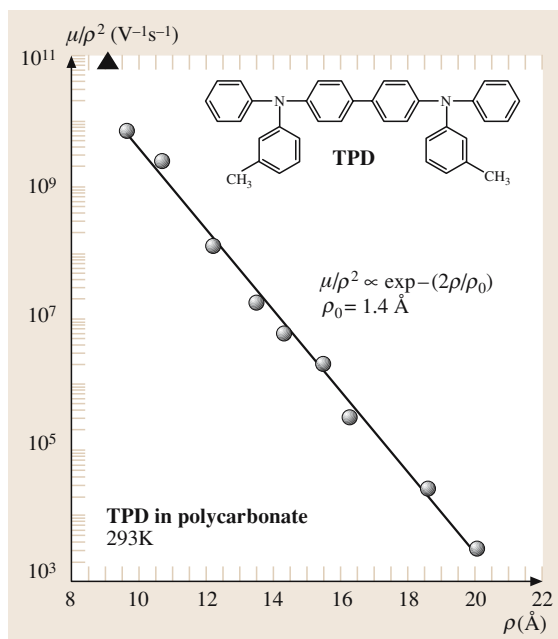


Fig. 39.2 Log (μ/ρ^2) versus average intersite separation, ρ , of TPD molecules in bisphenol-A polycarbonate ($E = 5 \times 10^4 \text{ V/cm}$). The triangle symbol is the hole drift mobility in pure TPD film

present in thermal equilibrium [39.28]. These systems can nevertheless support relatively efficient charge transport under the action of an applied field when in contact with a charge reservoir. At fixed temperature the drift mobility of extrinsic carriers decreases exponentially with increasing average intersite separation of the active molecule (ρ) so the drift mobility of MDPs can be tuned over a broad range by simply adjusting concentration [39.29, 30]. This feature of tunability can readily accommodate a wide range of xerographic process speeds as described above and is therefore advantageous from a technological point of view [39.31]. This is illustrated in Fig. 39.2 for a common hole-transport material, TPD N,N'-diphenyl-N,N'-bis(3-methylphenyl)-(1,1'-biphenyl)-4,4'-diamine (TPD, Fig. 39.4). The data is in conformity with a simple tunneling model where ρ_0 is the wavefunction localization radius and the transport states are sited on the TPD molecules. At 40–50 wt% doping approximately 10^4 hops are required for the hole to transit a 20- μm photoreceptor.

Drift mobility of holes in CTL films is most conveniently measured by the canonical small-signal time-of-flight (TOF) technique [39.32, 33]. However, for a given composition of the CTL, and all other conditions analogous, the same drift mobilities can be inferred by analyzing transport-limited xerographic discharge in bilayer photoreceptors [39.34, 35]. In TOF, the MDP film is prepared with semitransparent blocking contacts and maintained at a bias which is high enough to insure that the transit time of any excess injected carrier

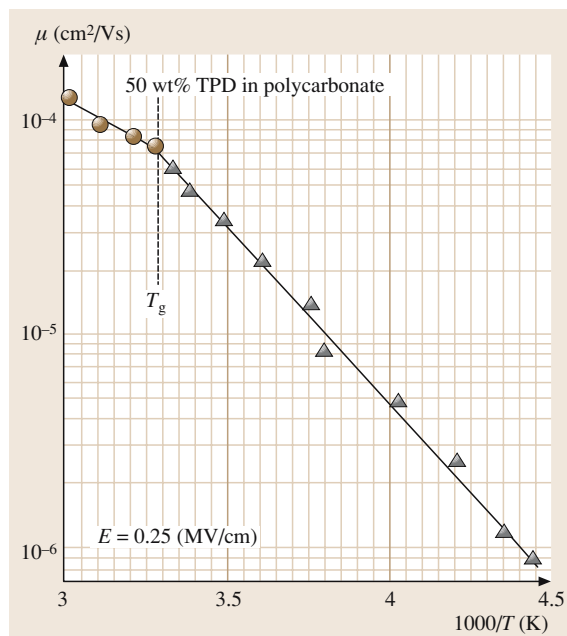


Fig. 39.3 Temperature dependence of the drift mobility in a film of 50 wt % TPD in bisphenol-A polycarbonate ($E = 2.5 \times 10^5$ V/cm). T_g is the glass transition temperature of the film

is shorter than the bulk dielectric relaxation time. The sample is exposed to a very short and weak pulse of strongly absorbed light (typically 337 nm from a ni-

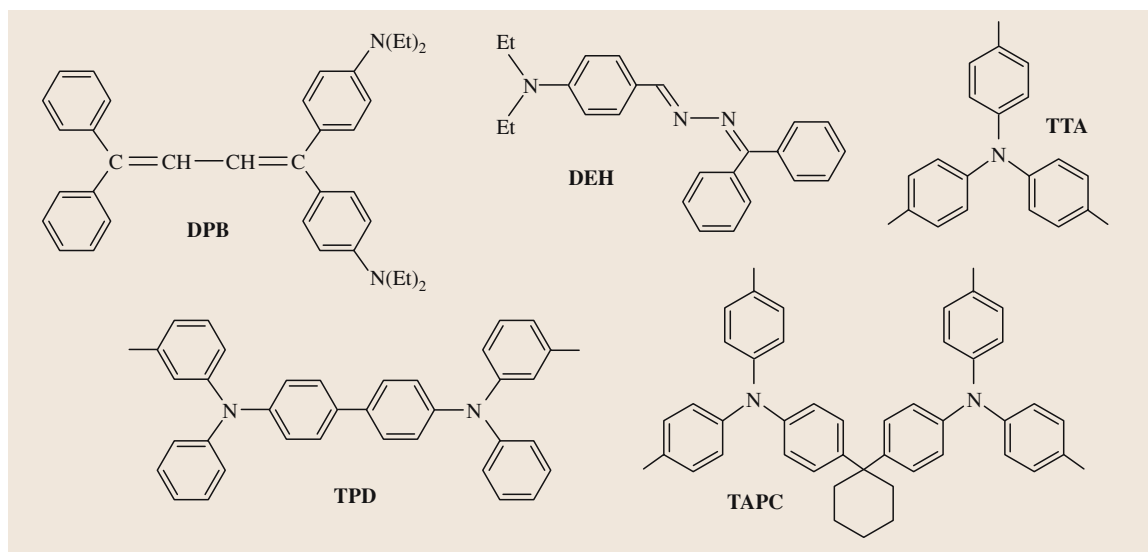


Fig. 39.4 Chemical structures of representative hole transport materials

trogen laser) incident on the positive electrode. Under these circumstances photoexcited transport molecules are oxidized at the positive electrode to the radical cation (the *hole* in transport terminology). This process sets up a chain of redox steps where electrons are progressively transferred from neutral molecules to their neighboring radical cation. The concentration of the advancing pulse of holes is low, such that the applied field remains uniform during their transit. In the ideal case the current from the advancing pulse of holes is constant until the leading edge reaches the counter-electrode, after which it rapidly decreases. The transit time can be extracted from the transient current and from this the mobility. For a given MDP composition drift mobilities are determined for films of known thickness as a function of applied field and temperature. The drift mobility in MDPs always has a thermally activated temperature dependence; thus the log of the drift mobility scales with inverse temperature, as illustrated in Fig. 39.3, where earlier measurements [39.29] have here [39.36] been extended to encompass the characteristic slope change displayed in the glass-transition region [39.37]. However, in a significant number of cases cited in the

literature [39.4] the analogous scaling is with the square of the inverse temperature, (i.e. non-simple activation), as predicted by the disorder model [39.38]. The experimentally observed scaling with intersite separation, the temperature dependence, and the high degree of disorder in these systems clearly indicate that electronic transport must involve the field-biased hopping of carriers in an energetically inequivalent manifold of states sited on the transport-active molecules. Apart from their technological importance in electrophotography [39.39] and light-emitting displays [39.40], what has made MDPs a laboratory for the study of hopping transport is that a number of key secondary features in their transport behavior are in fact pervasive in other disordered molecular media [39.41]. There is then the suggestion of a common underlying mechanism susceptible to theoretical treatment. Thus, the field- and temperature-dependent behavior of molecular dispersions is also observed in polymers with transport-active pendant groups such as poly(N-vinylcarbazole) [39.42], in polymers in which small molecule moieties are incorporated in a main

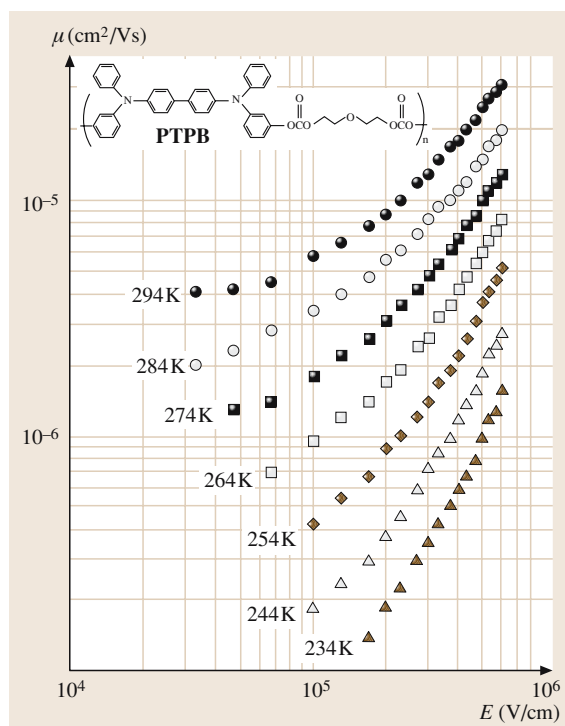


Fig. 39.5 Field dependence of the hole drift mobility in the polymer PTPB parametric in temperature

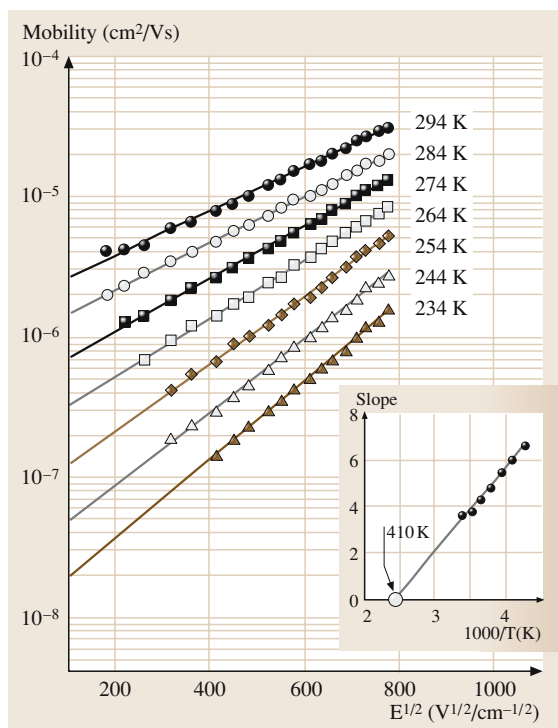


Fig. 39.6 Data set at each temperature in Fig. 39.5 replotted against $E^{1/2}$. The temperature dependence of the slope of each data set is plotted against reciprocal temperature in the insert

chain [39.43] in sigma-conjugated polymers [39.44, 45] and most recently in certain poly π -conjugated systems [39.46, 47]. Figure 39.5 is a plot of the field dependence of the hole drift mobility in the polymer polytetraphenylbenzidine (PTPB) (see insert for the molecular structure) versus field and parametric in temperature over a 60 K range [39.48]. PTPB is a glassy polymer in which tetraphenylbenzidine (TPD) moieties are covalently bonded into a main chain [39.43]. The data is replotted in Fig. 39.6 to show the explicit dependence of the log mobility on the square root of the applied field. The field dependence is itself temperature dependent and clearly becomes stronger with decreasing temperature. The two TPD-based hole-transport media are closely related but qualitatively identical results are reproduced in polysilylenes and polygermylenes which have a sigma-conjugated backbone capable of supporting electron delocalization as inferred from analysis of absorption and emission spectra [39.49, 50]. For example, hole transport in PTPB can be represented in an Arrhenius plot over the temperature range illustrated in Fig. 39.7. In this purely phenomenological

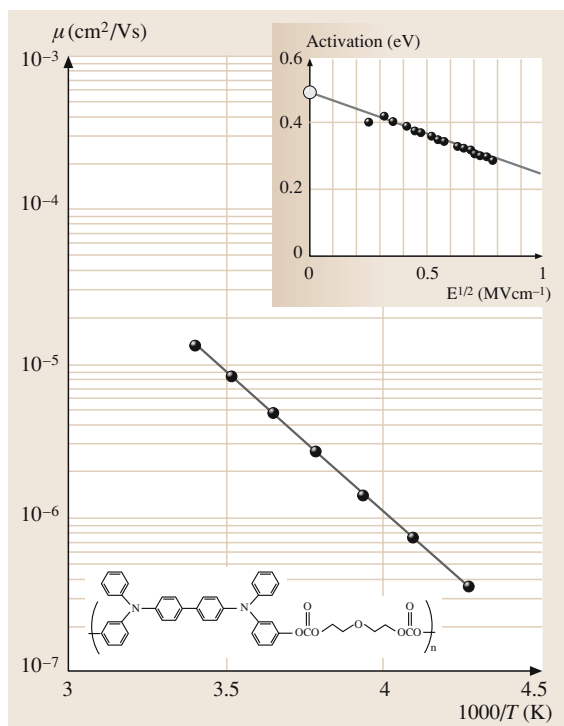


Fig. 39.7 Arrhenius representation of the temperature dependence of hole drift mobility in PTPB. The field dependence of activation is shown in the *insert*

description the activation decreases with the square root of the applied field, as displayed in the insert. Precisely the same behavior is exhibited when the same phenomenological description is applied to hole drift mobility data in poly(methylphenylsilylene) (PMPS), as shown in Fig. 39.8. Combining drift mobility [39.51] and spectroscopic data [39.52] has in fact suggested that transport in these sigma-conjugated polymers involves the hopping of holes among chromophore-like main-chain segments of varying length whose function for transport is therefore analogous to the TPD sites in PTPB. More recently the convoluted pattern of behavior first identified in MDPs and therefore clearly characteristic of hopping among discrete energetically inequivalent molecular sites has been reported in TOF experiments on π -conjugated systems notably the phenylene vinylenes. The sites, like those in polysilylenes, are interpreted to be domain-like backbone segments of varying conjugation length [39.46, 47]. These combined results and the recurrent pattern of behavior they reveal in systems

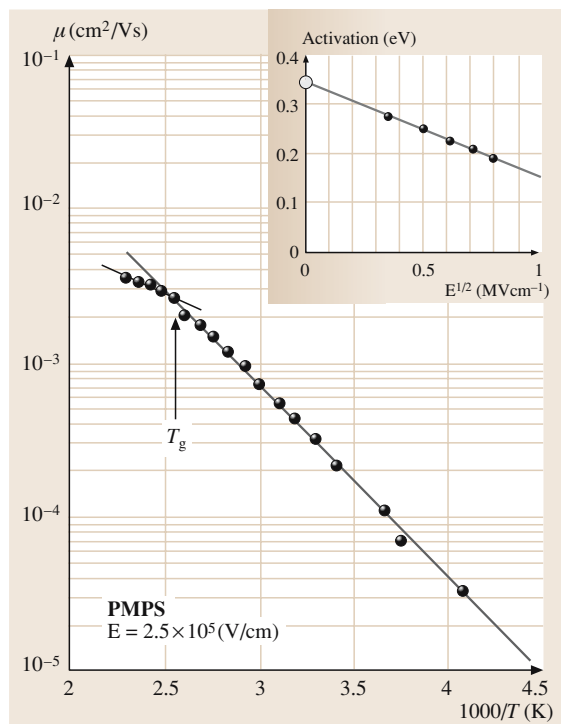


Fig. 39.8 Arrhenius representation of the temperature dependence of hole drift mobility in poly(methylphenylsilylene), PMPS. The field dependence of activation is shown in the insert. T_g is the glass transition temperature of the PMPS film

of widely varying composition and morphology suggest that a transport theory must be founded on universally shared characteristics [39.53]. The theoretical literature stimulated by these observations is in fact extensive and far too detailed to be adequately described here. In the present context it suffices to comment briefly on certain general trends. The activated temperature dependence universal to all these materials systems arises because states in the hopping manifold are energetically inequivalent. Energetic inequivalence can be understood on the basis of site relaxation accompanying polaron formation [39.54–56]. Energetic inequivalence of sites is also a common feature of disordered solids resulting from site-to-site fluctuation of the static and dynamic electrostatic potential [39.38, 57, 58]. The combination of both effects cannot be excluded. In fact, there is a general need to develop the framework to distinguish the relative contributions of these key processes in the analysis of experimental data [39.59, 60]. A particularly challenging issue has been the attempt to account for the special features of the Poole–Frenkel-like field dependence illustrated in Fig. 39.6. Polaron models typically predict that the log mobility should scale linearly with the applied field [39.54, 61]. The disorder model based on analysis of Monte Carlo computer simulation of hopping on a finite lattice as originally proposed by Bässler [39.38] was in fact able to self-consistently account for a number of key experimental features in terms of a limited set of disorder parameters. On the other hand, it could only model the observed field dependence over a very narrow range. It was later recognized in the formulation of dipole disorder models [39.62, 63], that slow site-to-site variation, as distinct from fully random fluctuation, in the effective electrostatic potential could more properly account for the field-dependent behavior of mobility commonly observed over several orders of magnitude [39.64–66]. The more general formulation of the disorder model developed by Dunlap and coworkers, treats the disorder potential as fully correlated and therefore slowly varying, and argues that such a supposition is physically plausible on general grounds in glassy solids.

If every chemical impurity remaining in even the most scrupulously processed polymer film could act as a trap, organics as a materials class would be completely excluded as a basis for the design of xerographic photoreceptors. Trap-free transport in CTLs was achieved not by entirely eliminating impurities but by properly designing the transport-active moiety [39.67, 68]. The associated studies first carried out in this technologically critical context on molecularly doped systems [39.69] were instrumental in generally

elucidating trap interactions in hopping systems and in first unambiguously illustrating and analyzing the mechanism of trap-controlled hopping transport [39.70]. It was already pointed out that hole transport is supported in MDPs when the neutral transport molecule is donor-like. For electrons the corresponding neutral transport molecule is acceptor-like in character. In each case, the CTLs are typically unipolar when analyzed by TOF or xerographic techniques. As it turns out the principle (simple in hindsight) which applies for example to hole-transport CTLs is to make the transport molecule significantly more donor-like than any of the resident impurity species. The limit is that the material must be stable to air oxidation. From an energetic point of view the impurity levels which then lie above the hole transport states are anti-traps. An important class of small molecules which have these desired characteristics are the aromatic amines and TPD is a prime example of a particularly strong donor in this class of molecules. The trap-free nature of these transport layers is xerographically apparent because there is no buildup of bulk space charge as measured by Kelvin-probe techniques [39.71] even after tens of thousands of CVs of holes are discharged through a photoreceptor with this CTL as a component. The trap-free nature of hole transport can also be independently demonstrated by combining TOF drift mobility and analysis of current–voltage measurements carried out on any transport layer fitted with a semitransparent blocking contact on one surface and an ohmic contact on the opposite face [39.72–74]. Figure 39.9 illustrates the result for the polymer PTPB in which TPD is covalently incorporated as the transport active unit [39.73]. Semitransparent Al on the films exposed surface is used for the TOF measurements of drift mobility. Dark injection occurs under positive bias from the gold-coated mica substrate onto which the PTPB film had initially been deposited. PTPB is highly insulating so that the bulk dielectric relaxation time is always much longer than the transit time of any excess injected carrier at all applied voltages. With the Al contact under positive bias, hole drift mobility is measured as a function of applied voltage V . From the measured drift mobility the trap-free space-charge-limited current density (TFSCLC), J , [39.75] that would be sustained by an ohmic contact under positive bias can be calculated to a good approximation, even when the drift mobility is itself field dependent [39.76, 77] using the expression $J = 9\epsilon\epsilon_0\mu V^2/8L^3$. These values are represented by open circles in Fig. 39.9. The electric field in the film is, of course, nonuniform and E in plots of this sort is simply a shorthand for V/L ,

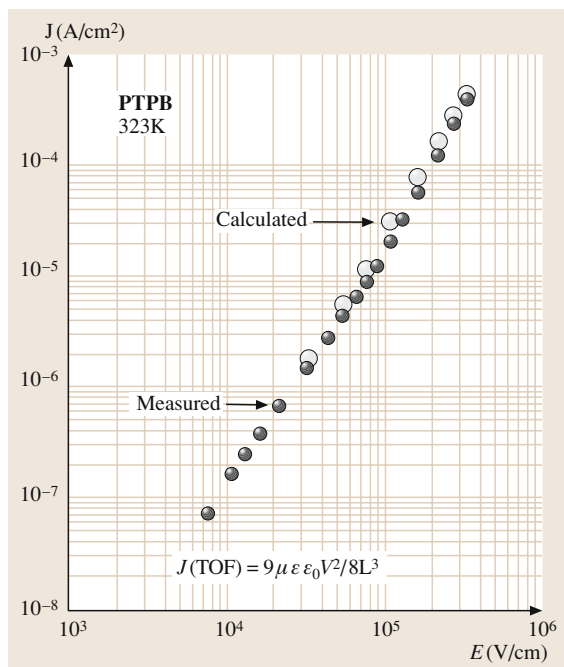


Fig. 39.9 The field dependence of the dark current for a PTPB film at 323 K measured with positive bias applied to an Au coated Mica substrate (*filled circles*). The trap free space charge limited current (*open circles*) is calculated from the hole drift mobilities obtained by the time of flight method

where V is the externally applied voltage. It should be noted that, under the present conditions, the TFS-CLC is in principle the maximum current that the bulk of a transport medium such as PTPB can demand from a contact in the steady state. It should also be noted that the observation that an injecting contact under test is capable of sustaining a space-charge-limited dark current is a *prima facie* demonstration of ohmicity [39.24]. Filled circles represent the steady-state current measured when positive bias is applied to the substrate. That these calculated and measured currents coincide demonstrates that the injecting Au contact is ohmic and that the PTPB film is trap-free. Similar results have been obtained for other commonly used CTL MDPs such as TPD in bisphenol-A polycarbonate and in films of poly(methylphenylsilylene).

39.2.3 Photogeneration

The CGL material is designed to strongly absorb light from the exposure system (red to near-infrared LEDs

and lasers are typical). In a typical dual-layer OPC the CGL material is a thin (submicron) coating of a dispersed pigment. Simply, the pigment material absorbs a photon and mobile charge is created which is *injected* into the CTL and then drifts under the influence of the applied field to discharge the device. Charge generation is a very complex topic [39.78]. In general terms, absorption of a photon produces an excited state (exciton) of a correlated hole–electron pair (valence band and conduction band, respectively) which has some mobility. Measured exciton ranges can be on the order of 0.1 μm . The exciton may then undergo recombination, with the release of energy to the environment, or it may be converted into charges. Conversion into charges can occur *spontaneously* or through chemical (impurities or dopants) or physical (crystal defects or the crystal surface) interactions. With the trapping of one of the charge carriers the other is free to drift under the influence of the applied field and geminate recombination is suppressed. The mobile carrier (assume holes) must drift to where it is in contact with the hole transport moieties of the CTL. In fact, in most cases the CTL penetrates and mixes with the CGL so that, even if the CGL is formulated without transport material, we can assume that these molecules are in molecular contact with the pigment particle surface. Now the *hole* is available to oxidize or *inject* into the transport moiety to produce the radical cation. Useful hole-transport materials have relatively low oxidation potentials [≈ 1 V versus the saturated calomel electrode (SCE)]. The oxidation potential must be lower than that of the hole at the pigment surface but not so low that the material is unstable to air oxidation. The CTL hole is now transported through the CTL, discharging the photoreceptor, by sequential one-electron redox reactions between adjacent transport molecules as described in the previous section.

The charge-generation mechanism is characterized as *intrinsic* when the hole is generated within the pigment particle without the influence of any externally added materials. In this case the hole is injected into the CTL with concomitant oxidation of a transport moiety at the pigment surface. *Sensitized* generation is said to take place when the photoconductivity is enhanced by an added material (dopant). In this case the dopant is available to react with the pigment exciton at the particle surface. In most cases the dopant is in fact the hole-transport moiety. Dopants may also influence charge generation by forming a (charge-transfer or other) complex which facilitates charge separation. Such complexes may be accessed by direct photoexcitation

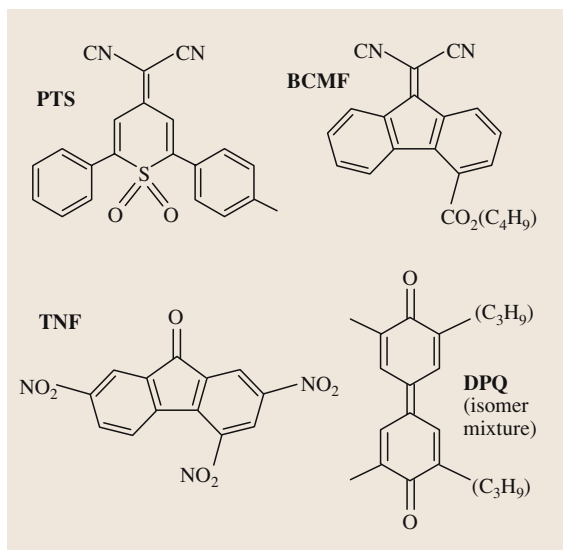


Fig. 39.10 Chemical structures of representative electron transport materials

or via exciton diffusion. Sensitization of the photoconductivity of poly(*N*-vinyl carbazole) (**PVK**) by electron acceptors such as **TNF** (Fig. 39.10) has been extensively studied [39.79–87]. Charge-generation mechanisms in pigment-based CGLs depend on the pigment molecular structure, crystal structure, morphology, and on the transport molecule used in the formulation.

The fundamental quantity of interest in charge generation is the quantum efficiency as a function of field. A strong field dependence is always observed, as shown in the example of Fig. 39.11. Several models have been proposed to explain charge-generation characteristics. These will be summarized here but more complete discussions can be found in *Pope and Swenberg* [39.78], *Popovic* [39.89], *Law* [39.3]; *Borsenberger and Weiss* [39.4], and *Weiss et al.* [39.6]. One of the early models to be applied to organic photoconductors was the Onsager model of geminate recombination [39.90]. The 1938 theory gives the probability that a charge pair in thermal equilibrium will separate under the influence of an applied field in competition with geminate recombination. In this model charge separation occurs from an upper vibrational state to produce a *geminate* pair of charges which either recombine due to their Coulombic attraction or are separated under the influence of the applied field. The parameters in this model are the field-dependent quantum efficiency for geminate-pair formation, η_0 , and the initial separation distance of the geminate pair (r_0) [39.91]. The data in Fig. 39.11

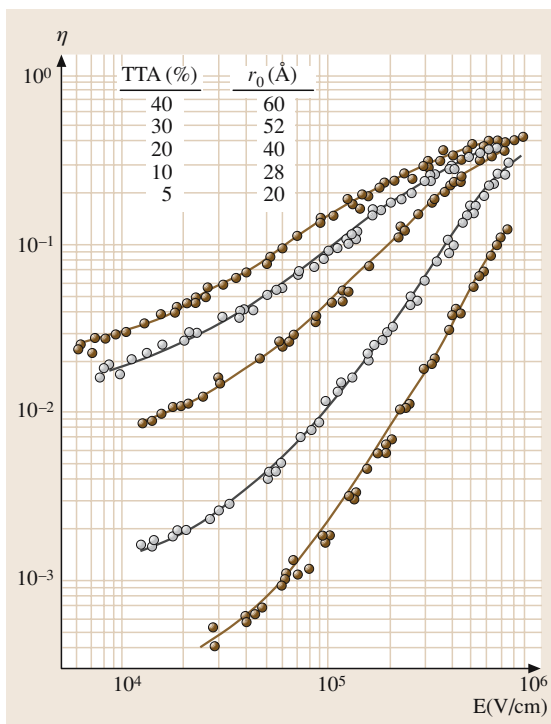


Fig. 39.11 Quantum efficiency of the field dependence of generation for a dual layer dye-polymer aggregate based OPC. The solid lines were calculated from the Onsager model with $\eta_0 = 0.60$ and r_0 as shown. After [39.88]

is for a dual-layer dye-polymer-aggregate-based OPC with varying concentrations of charge-transport material (tri-*p*-tolylamine, TTA) in the CTL [39.88]. The data has been fit to the 1938 model with $\eta_0 = 0.60$ and r_0 as shown. The model has been refined using a modified basis, also due to *Onsager* [39.92]. Further refinements have involved modifications of the distribution function assumed for the separation distances [39.93, 94]. This model has been successfully used to describe charge generation in amorphous materials [39.95]. However, these models predict a wavelength-dependent quantum efficiency and a photoconduction threshold which is at higher energy than the absorption, neither of which is observed in pigment-based generation materials. A model for charge generation in these materials invokes decay (radiative or radiationless) of the initially formed excited state followed by a competition between field-dependent charge separation and internal conversion [39.96]. Sensitized generation occurs with the reaction of the thermalized excited state with a dopant. Studies of the field dependence of pigment

fluorescence quenching and lifetime have been used to sort out the details with specific generation materials [39.89,97–99].

The quantum efficiency for carrier generation can be defined in several ways, as described above. The efficiency of geminate-pair formation refers to the initially produced charge pair and this is also the efficiency which has been obtained with fluorescence quenching studies. On the other hand, the generation efficiency determined with xerographic discharge al-

ways includes some contribution from charge transport because a measured decrease in photoreceptor surface potential is due to charge motion. In addition, very-high-intensity exposures may give rise to (bimolecular) recombination of positive and negative carriers during transport [39.100–104]. Quantum-efficiency determinations from xerographic discharge are carried out with very short, low-light-intensity continuous exposures to minimize the influence of charge transport and recombination (see the following section).

39.3 OPC Characterization

Photoreceptors may be characterized by many techniques. On the one hand one wishes to determine the fundamental characteristics of the materials, while on the other hand one wishes to determine electrophotographic performance-based characteristics of the photoreceptor. The quantum efficiency of charge generation and, depending on the model chosen, the initial separation distance of the charge pair, are common characterizations. Electrophotographic performance characterizations include dark decay, photoinduced discharge sensitivity, and residual potential as well as their stability to electrical-only cycling.

39.3.1 Dark Decay

The quantity of interest is the rate of dark discharge (V/s) as a function of field and environment. However, the dark decay characteristics of a dark-adapted photoreceptor when incrementally corona charged to higher and higher fields will be significantly different from what is observed in the electrophotographic process. This is because OPCs seldom behave as perfect capacitors and there is typically a depletion charging component where previously trapped charges are mobilized and drift in the photoreceptor on the application of the corona charge [39.105]. Thus, in addition to field and environment, the observed dark decay characteristics depend on prior exposure history (trapped charges), corona charger characteristics, and timing. Thus, a given photoreceptor will have a higher dark decay rate if the photoreceptor is charged exactly to the desired initial potential (V_0) and the dark decay rate determined immediately, as opposed to charging well above the desired potential and measuring when the surface potential decreases to the desired potential. Dark decay measurements which are meaning-

ful to the electrophotographic process must be carried out in an apparatus in which the exposures, charging, and timing elements are fixed. Dark decay is an important characteristic of an OPC because, although process control will attempt to maintain V_0 , the imaging system in a printer will in general not tolerate a significant V_0 drop.

39.3.2 Photosensitivity

Measurements to determine fundamental characteristics (quantum efficiency of generation materials, mobility, etc.) are often carried on photoreceptors that differ from what might be eventually used in a printer. The layer of interest (CGL or CTL) may be isolated or the device may be modified with the application of a vacuum-deposited electrode (Au is common) on the free surface. Common fundamental characterizations include quantum efficiency as a function of field and wavelength, and mobility. Mobility was discussed in Sect. 39.2.2. The quantum efficiency of carrier generation has several definitions as applied to OPCs. On the one hand it can refer to the number of ion pairs generated per incident or absorbed photon. This fundamental characteristic might be determined experimentally by fluorescence quenching, or calculated based on a model such as the Onsager model discussed above. On the other hand a photodischarge efficiency can be determined based on the decrease in surface potential per photon absorbed or incident. The photoinduced discharge (PIDC) method is commonly used to characterize OPCs as a function of field, wavelength, temperature, humidity, exposure intensity, etc. The photoreceptor is corona charged to apply the desired field and is photodischarged. Using the parallel-plate-capacitor model, $Q = CV$ (terms as defined previously), the change in surface charge den-

sity can be calculated from the decrease in the surface potential. The change in surface charge density per unit area per photon incident or absorbed is a measure of photosensitivity. This measure includes the efficiency of charge generation, injection into the CTL, and transport through the CTL. Because of carrier bimolecular recombination and range limitations the quantum efficiency determined using this method may depend on the light intensity and the extent of photodischarge. The photodischarge method is carried out with either continuous or flash exposures. With continuous exposures ($\text{J}/\text{cm}^2/\text{s}$) the surface potential is monitored as a function of time. One measure of OPC photosensitivity which relates to the electrophotographic process is the exposure required to discharge the photoreceptor a fixed amount, often 50%, from an initial potential. This is typically carried out using exposures (wavelength and duration) relevant to a particular electrophotographic process. The photoreceptor sensitivity is usually reported in terms of J/cm^2 (the inverse is often used to report spectral sensitivity). A related method is to determine the initial photodischarge rate (dV/dt). This can be related to the decrease in surface charge density through the capacitance, $dQ/dt = C dV/dt$. The light intensity is kept low and the discharge time short to avoid space-charge perturbations during charge transport and to keep the field essentially constant. The ratio of the decrease in surface charge density to the exposure ($\text{photons}/\text{cm}^2$) is a measure of photodischarge efficiency. In fact the photodischarge per exposure ($\text{V}/\text{J}/\text{cm}^2$) is another commonly used metric for OPC photosensitivity. When determined as a function of field, data obtained in this manner can be analyzed in terms of the Onsager models. Because photosensitivity depends on field (surface potential per unit OPC thickness) and the exposure produced decrease in surface charge density (related to the change in surface potential through $\Delta Q = C \Delta V$), meaningful comparisons can only be made between OPCs of similar thickness, charged to similar initial potentials, and exposed to similar discharged potentials.

For a process using a flash exposure the photoreceptor discharge is determined at several exposure levels (with the surface potential determined at a fixed time after the exposure) and the data is plotted as surface potential (V) versus $\log(\text{exposure}, \text{J}/\text{cm}^2)$. An arbitrary photodischarge point (typically 50%) can be determined

from this plot for comparison purposes. The utility of this characterization is that the exposure characteristics and the timing for reading the discharged surface potential directly relate to the printing process. As discussed above, short high-intensity exposures may give rise to Langevin recombination [39.102, 104, 106], resulting in decreased sensitivity (reciprocity failure) relative to low-intensity exposures.

An often overlooked sensitivity factor is that there are more photons per energy unit (J) as the wavelength increases. For example, there are about 30% more photons/ J at 820 nm relative to 630 nm. Thus, for a photoreceptor with a given quantum efficiency for carrier generation the sensitivity will increase with increasing wavelength of the exposure source.

39.3.3 Electrical-Only Cycling

Electrophotographic performance characterizations are necessarily carried out on the photoreceptor in an apparatus which has the key elements of the electrophotographic process. The most *relevant* apparatus would of course be the fully configured printer. However, the use of such a complex device is often undesirable, or unnecessarily, for screening purposes. In the laboratory it is common to use *electrical-only* test fixtures in which the process always includes charge and expose, and may include erase (or unique process elements), with the surface potential being determined after the application of each. With an apparatus such as this it is possible to determine process-relevant characteristics such as: chargeability, dark decay, photosensitivity, and residual potential, all as a function of cycling. Of course large changes in any of these characteristics is generally undesirable and long cycling (10–100 kcycles) is often necessary to adequately determine the electrical stability of the photoreceptor to the electrophotographic process. Such stability is a necessary, but not sufficient, condition for eventual commercialization, and extensive testing in the machine for which the photoreceptor was designed is always the final step in the development process. Other factors, often connected to the device physical characteristics (wear rate, corona chemical sensitivity, layer adhesion, seam strength, etc.) may in fact determine the eventual success or failure of a particular OPC in a printer.

39.4 OPC Architecture and Composition

As we have described, OPCs are large-area thin-film devices that are insulating in the dark and become conductive when exposed to light. A high field is applied by exposure to an air corona (or similar charging method) and the photoreceptor is image-wise exposed to produce an electrostatic latent image. The surface potential in the exposed areas is determined by the exposure intensity and the OPC photosensitivity while the surface potential in the unexposed areas is determined by the dark conductivity. The development system deposits charged marking particles (toner) on the photoreceptor surface according to the difference in surface potential between the exposed and unexposed areas. The key photoelectrical processes which enable latent image formation are photoinduced charge generation and the transport of the generated charge through the device. Organic photoreceptors accomplish this in a multilayer architecture with specific materials (monomers and polymers) designed and optimized for specialized functions: binder, charge generation, charge transport, dark conductivity, etc. In this section we will discuss the architecture and composition of organic photoreceptors with details on the function and composition of each layer.

39.4.1 OPC Architecture

The choice of photoreceptor thickness is a balance between capacitance, surface potential, and internal field. Capacitance relates to surface charge density, which determines the achievable developed toner density for toners with a given surface charge per area. Surface potential, in combination with the development system, provides the potential difference to attract toner to the photoreceptor surface. The internal field enables efficient charge generation and transport. In most electrophotographic printing machines the balance is achieved with an OPC thickness of $\approx 20\text{--}30\text{ }\mu\text{m}$ and a surface potential of 500–1000 V.

A schematic cross section of a fully configured modern electrophotographic photoreceptor is shown in Fig. 39.12. In its most simple form the photoreceptor consists of a conductive substrate and a single layer of photoconductive material to accomplish charge generation and transport. The advantage of this structure is in its simplicity, but the disadvantage is that not all characteristics will be optimum. In the modern embodiment the photoreceptor is constructed with separate layers for charge generation (CGL) and charge transport (CTL). Other layers shown in Fig. 39.12 include: sub-

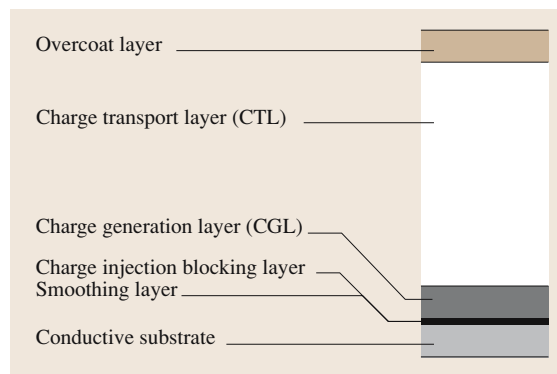


Fig. 39.12 A schematic cross section of a fully configured “modern” electrophotographic photoreceptor

strate, smoothing layer, charge-injection blocking layer, charge-generation layer (CGL), charge-transport layer (CTL), and protective overcoat layer. Additional layers, such as adhesive interlayers, and a back-side anti-curl layer may also be present. OPCs with separate CGL and CTL are termed dual-layer. Each of these layers may be a complex formulation of materials. The advantage in separating layers is that each layer can be optimized for its specialized function. This adds materials expense and manufacturing complexity but has the advantage of superior performance.

In a single-layer photoreceptor when exposed with strongly absorbed radiation directed at the free surface, charge generation will occur near the free surface and the photodischarge will be dominated by the transport of positive or negative charge through the bulk of the layer. The polarity of the mobile charge will be the same as the applied surface potential. In most commercial dual-layer photoreceptors the architecture is substrate–CGL–CTL. Because the CTL is formulated for hole transport, dual-layer OPCs are charged with a negative surface potential. Positive corona charging has some advantages over negative (better uniformity and less ozone production). A single-layer OPC can function with either a positive or negative surface charge depending on the exposure direction and the transport characteristics of the materials used in the formulation. Positive-charging dual-layer photoreceptors with the architecture substrate–CTL–CGL have been commercialized but these require a protective overcoat to prevent the CGL from being damaged by contact with the electrophotographic process elements. The typical substrate–CGL–CTL architecture will function with

positive charging if the CTL is formulated for electron transport. Electron-transporting CTLs have been widely studied by not yet commercialized.

In a multilayer photoreceptor the principle is that the functions of each layer are optimized. Thus, with separate CGL and CTL the former is designed for strong absorption of the wavelength of light used in the process and efficient charge generation, and the latter is optimized for the injection of photogenerated charge from the CGL and transport of that charge to the free surface. In general, most charge-generation materials efficiently transport only one polarity of charge (electrons or holes depending on the material) so that it is generally desirable to have the CGL as thin as possible while retaining the required optical density at the exposure wavelength. It is also necessary that the CGL be uniform over the entire area of the photoreceptor to maintain consistent sensitivity for the entire photoreceptor surface. Submicron CGLs are typical of the photoreceptors used in most modern digital printers. The CTL determines the device capacitance and a thickness of $\approx 25\text{ }\mu\text{m}$ is typical. However, the CTL thickness also influences the ultimate latent-image resolution in a digital printer because holes transiting the CTL experience mutual Coulombic repulsion [39.12, 107]. It has been shown that for optimum resolution the photoreceptor thickness should be about one quarter of the pixel size [39.12, 13]. Thus, for 600-dpi imaging ($\approx 40\text{ }\mu\text{m}$ pixel size) the ideal photoreceptor is on the order of $10\text{ }\mu\text{m}$. Thin OPCs have not yet been introduced into digital printers but one might anticipate that this will occur as toners get smaller and the processes of toner transfer and fusing are refined to minimize image disruption [39.108].

39.4.2 Coating Technologies

Hopper coating of continuous webs and dip coating of drums are the two technologies commonly used to fabricate photoreceptors. As with any large-area photoelectrical device absolute cleanliness is required to prevent artifacts in the coated layers. In a multilayer device one must carefully control the interfaces between the layers. For example, a key interface is that between the CGL and CTL. Here it is important that the two layers adhere but do not mix to such a degree that the desired characteristics of either layer are compromised. Attaining the desired degree of mixing is a process of optimization depending on the materials, solvents, and coating method.

In hopper coating the coating solution is continuously pumped through a slot onto the moving web.

Film thickness is controlled by the pumping rate and solids content of the coating solution. Precise control of drying temperature, humidity, and air flow enable optimization of the coated layers. Curing of the coated film involves removal of solvent and annealing of the polymer binder(s). In a hopper coating process several layers can be coated in a single pass with sequential hoppers as long as the previous layer is adequately cured. Hopper coatings can be carried out at a very high speed (m/s) to produce a roll of many thousands of feet. To coat subsequent layers the previously coated substrate is re-run through the coating machine. Because the hopper coating process involves contact of the coated surfaces, and winding of the coated web, there is the possibility of film damage. Final inspection for film defects must be carried out and damaged, or imperfectly coated, regions must be removed before construction of the final photoreceptor loops.

In a dip coating process the substrate and coated layers are dipped into a solution or dispersion of the next layer to be coated and slowly withdrawn and cured. The coated layer thickness is determined by the rheological characteristics of the coating solution and the rate of withdrawal. Under these conditions there is the opportunity for extensive interaction between the previously coated layer and the solution into which it is dipped. Thus, formulations which are suitable for hopper coating may be inadequate for dip coating even though the chemical compositions of the final coated layers might be identical after curing. Another consideration with dip coating is that components of the previously coated layers which are soluble in the solvent used in the subsequent dip will contaminate the dip reservoir. Low levels of contamination may be insignificant, but since the amount of fluid removed in each coating operation is a small fraction of the total volume one must make sure that such contamination does not degrade performance.

Other processes such as ring coating (a dipped drum is withdrawn through an annulus which acts as a doctor blade to control the wet coating thickness), spray coating, vacuum coating, etc. may be used for specific applications and materials. In all layers coated from solution it is often necessary to add a small quantity of surface-active agent as a coating aid to enable smooth, defect-free coatings.

In the following sections we will discuss details around the compositions of each of the layers which might be found in an organic photoreceptor for electrophotographic printing. The order of discussion is in the order in which the typical photoreceptor would be constructed: substrate and ground layer, smoothing

layer and charge-blocking layer, charge-generation layer (CGL), charge-transport layer (CTL), and protective overcoat layer.

39.4.3 Substrate and Conductive Layer

A common element to all electrophotographic photoreceptors is a substrate on which all the layers are deposited. In a research environment a small square of glass or quartz is convenient, but in commercial applications the substrate is invariably either a polymer film or a metal drum. Some of the material characteristics which might be considered in the choice of a polymer film are: stiffness, toughness, transparency, core set, and surface friction. Polymeric photoreceptor substrates are often 3–7 mil. A common polymer film material is the polyester poly(ethylene terephthalate) (PET). PET film is very durable, optically transparent, and readily available in a wide range of thickness. Since it is highly insulating it must be coated with a conductive material to act as the ground plane. PET or other polymer films are available in large rolls and the coating operations are usually done on the roll rather than cut-to-size sheets. Although a solution-coated conductive layer is possible, the most common ground layers are metals such as Ni, Al, Cr, and Ti, which are applied in a continuous sputtering or similar vacuum technique. The metallized rolls are subsequently coated with the photoconductive layers.

If a metal drum is the chosen substrate it is almost invariably aluminium (1–10 mm wall thickness). The conductive metal serves as the ground layer but typically it receives further treatment before coating the photoconductive layers. The treatment may be chemical such as anodization to form a thin hole-blocking layer, or physical such as turning or sanding. Because these secondary surface treatments add cost, a smoothing layer is often used instead.

The electrical conductivity requirements of the conductive layer are modest. The higher the process speed the greater the conductivity required for adequate surface charge accumulation. Analysis by Chen [39.109] demonstrated that the result of inadequate conductivity in corona charging of a photoreceptor is a nonuniform surface potential. The surface potential is decreased depending on the distance from the actual grounding contact. For a typical situation the upper limit of the sheet resistance is $\approx 10^4 \Omega$ per square.

Alternative approaches such as *seamless* webs and conductive plastic drums have been developed but have not yet been commercialized.

39.4.4 Smoothing Layer and Charge-Blocking Layer

For xerographic applications it is important that the conductive layer does not inject charge into the photoreceptor in the dark. However, the metals used as grounding layers all have the potential to inject holes into the photoreceptor CGL via direct contact with the charge-generation material or the hole-transport material. The injection and transport of charge in the dark would cause a loss of surface potential (dark decay). Since the formation of a latent image relies on a difference between the rate of dark- and light-induced surface potential decrease, a high dark decay rate means that the process must be amenable to a small imaging potential. Another aspect of dark injection from the conductive layer is that such phenomena often occur in small localized *charge-deficient spots*. In a process where the discharged areas are developed small deposits of toner are found in what should be non-toned background. To counter such processes it is common to interpose a charge- (typically hole-) blocking layer between the ground layer and the CGL. With drum photoreceptors it may be the roughness of the metal surface which initiates the charge injection. In this case it may be necessary to interpose a relatively thick *smoothing layer* between the metal surface and the CGL.

In principle, any good film-forming insulating material will suffice as a charge-blocking layer as long as adhesion to the ground layer and CGL are adequate. Several types of materials have been used for this purpose including polysiloxanes and nylons. Since these materials are insulating, in an electrophotographic charge/expose cycle there will be a residual potential due to the field remaining across the layer. This will build up with cycling depending on the layer conductivity and the process cycle time. Although residual potentials typically decrease when the photoreceptor is rested in the dark, in practice such relaxation is usually incomplete and in subsequent cycling the residual builds up at a faster rate. The physics and chemistry of such hysteresis is not well understood. At some point the residual potential may cause image degradation and necessitate photoreceptor replacement. Thus, the blocking layer must be thick enough to cover the ground layer uniformly but not so thick as to cause an undesirable residual potential. Generally such a layer is coated at around $1 \mu\text{m}$ or less.

With an aluminium drum a thin anodized layer may prevent charge injection. However, making the drum surface smooth on a submicron scale is expensive

and an alternative is to use a relatively thick smoothing/blocking layer to cover the surface irregularities. Smoothing layers are typically metal oxides (titanium for example) in a polymer binder. The oxide loading is high to impart sufficient conductivity to prevent a residual potential. If the metal oxide does not inject holes it also serves as the charge-blocking layer. Such layers are up to around 10 μm thick.

Imaging with monochromatic radiation in digital printers has led to a new imaging artifact caused by interference between the incoming and reflected light. In dual-layer OPCs with the CTL uppermost, if the CGL incompletely absorbs the exposing light it can be reflected from the ground layer. The interference pattern that is produced can be visible in the developed image as a *wood grain* pattern. One way to counter this is to scatter the incoming light before it's reflected. There are many patents describing techniques for carrying this out including roughening of the surface of the conductive layer.

39.4.5 Charge-Generation Layer (CGL)

The CGL is where the imaging light is absorbed and charges are generated. The first major consideration in choice of material is that it strongly absorb the wavelengths of light emitted by the exposure system. In copiers light is reflected from a document onto the charged photoreceptor. Since an original may have colored components the light source must cover the visible range (400–700 nm) and the photoreceptor must have broad sensitivity to visible light. Common light sources in copiers are fluorescent or xenon lamps. Both sources are generally filtered to remove ultraviolet and infrared light. In digital printers the exposing source may be an LED array (680–780 nm) or a scanned laser (633 nm from He–Ne or 740, 780, and 820 nm from laser diodes).

Most modern commercial photoreceptors utilize pigments as charge-generation materials. Pigment-based CGLs comprise submicron crystals of the charge-generation material suspended in a polymer binder. Pigments are milled to the desired size and stabilized with appropriate additives and binders to produce a coating dispersion. Charge-generation materials are generally classified according to their chemical identity. Materials which absorb primarily in the visible region are dye-polymer aggregate, PVK–TNF charge-transfer

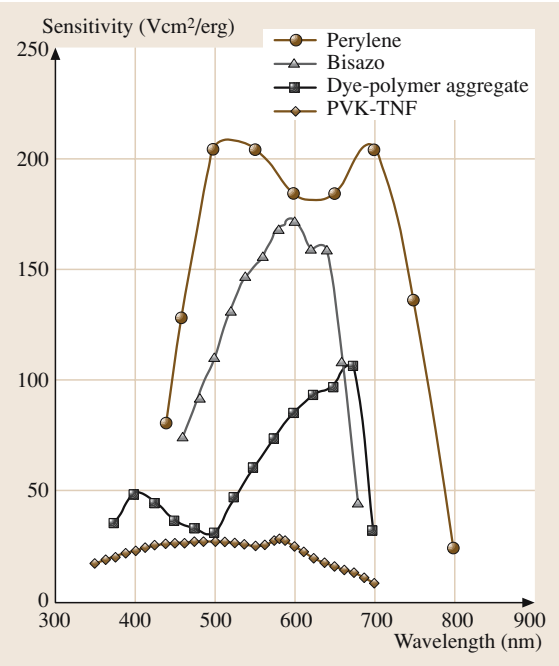


Fig. 39.13 Sensitivity action spectra of visible sensitive OPCs. The chemical structures of the CGL materials are given in Fig. 39.15

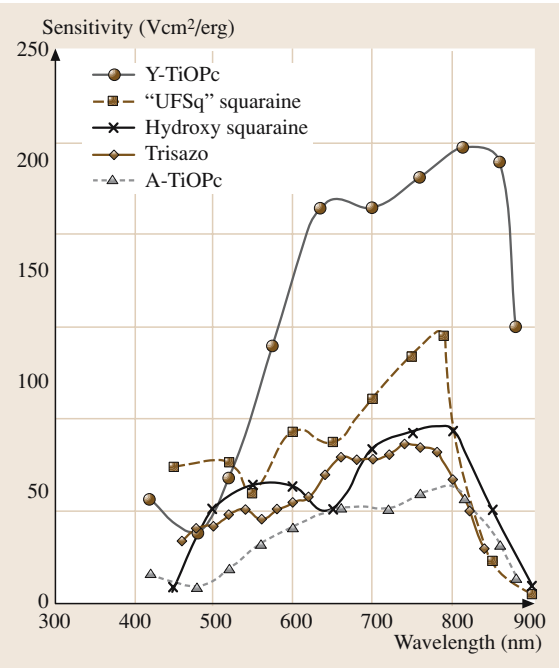


Fig. 39.14 Sensitivity action spectra of near-infrared sensitive OPCs. The chemical structures of the CGL materials are given in Fig. 39.16

complex, perylenes, and bisazos. Phthalocyanines, trisazo, and squarylium pigments are utilized for near-infrared absorption. Figures 39.13 and 39.14 show the sensitivity action spectra and Figs. 39.15 and 39.16 the chemical structures of these charge-generation materials. The sensitivity action spectra in these figures are taken from literature data and the sensitivities converted into consistent units ($\text{V}/\text{cm}^2/\text{erg}$) based on information given in the cited publication. Because the data are from OPCs with different thickness, exposure intensity, exposure time, initial potential, and exposed potential, the sensitivities are not directly comparable. However, the wavelength ranges are accurate. When the exposure is through the CTL (in a negative-charging dual-layer OPC) the photosensitivity will be decreased at wavelengths below $\approx 400\text{ nm}$ because of light absorption by the transport material.

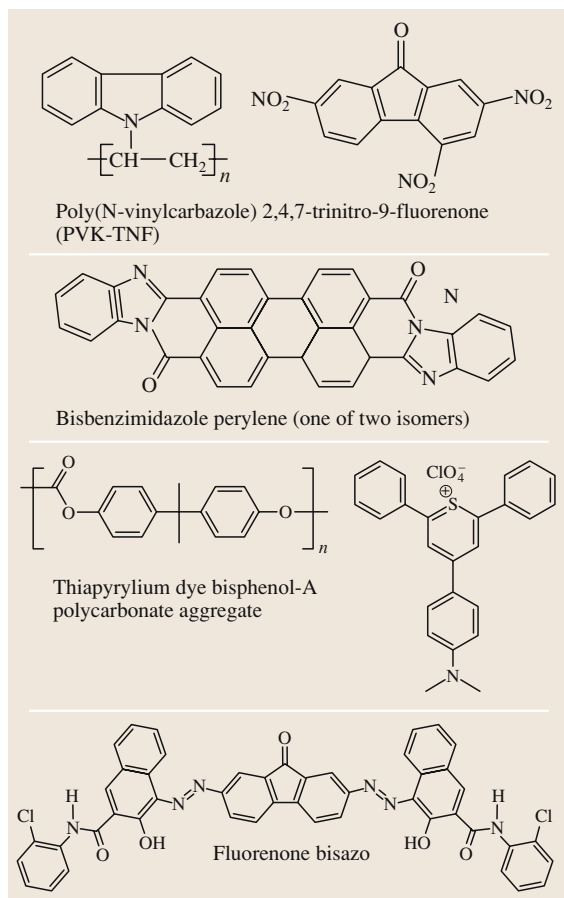


Fig. 39.15 Chemical structures of visible sensitive charge generation materials

Perylenes, azos, and phthalocyanines are best known for their use as colorants and it is because these materials were readily available that they were investigated as charge-generation materials. A complication with pigments is that some exhibit polymorphism. Over many decades the colorant industry has developed technologies for controlling polymorphism to obtain pigments with specific colorant characteristics and similar technologies have been recently developed to obtain

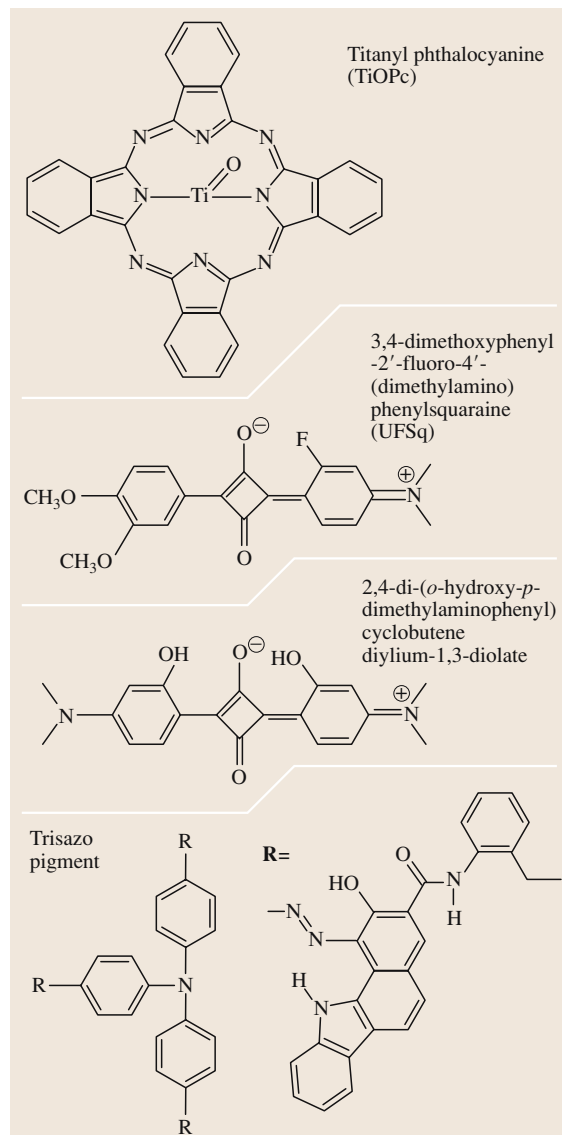


Fig. 39.16 Chemical structures of near-infrared sensitive charge generation materials

polymorphs with the appropriate absorption as well as charge-generation characteristics for electrophotographic applications. Thus, in many cases to prepare a CGL it is first necessary to convert the pigment-generation material into a dispersion of submicron particles of the desired polymorph. Examples of charge-generation materials are discussed below.

The donor-acceptor charge-transfer (CT) complex between poly(N-vinyl carbazole) (PVK) and 2,4,7-trinitrofluorene-9-one (TNF) was the first commercialized organic photoreceptor [39.110, 111]. The photoconductivity characteristics have been extensively studied [39.79–81, 84–86] and reviewed [39.112–115]. With the addition of TNF the carrier-generation efficiency of PVK is increased and extended into the region of CT complex absorption. Based on field quenching of exciplex fluorescence a mechanism for PVK-acceptor systems was proposed involving carrier generation from a nonrelaxed exciplex which is accessed by either direct absorption into the CT state or via an encounter complex between excited PVK and TNF [39.84–87]. As used in the IBM copiers I and II a 1 : 1 molar ratio of PVK to TNF was coated as a single 20- μm layer and electron-dominated transport was used to create the latent image (negative surface charge) with a He-Ne laser exposure (632.8 nm) [39.111]. An exposure of 5.5 $\mu\text{J}/\text{cm}^2$ was used to decrease the surface potential from -750 to -200 V. The quantum efficiency based on incident photons was ≈ 0.2 while the maximum efficiency was ≈ 0.3 at ≈ 400 nm. The photosensitivity action spectrum for this OPC is shown in Fig. 39.11 [39.110]. Using the photoinduced discharge technique [39.82] the maximum quantum efficiency (charges transported through the device per photon absorbed) was found to be 0.14 at 375 nm and a field of 10^6 V/cm. The PVK-TNF-based OPC had several drawbacks: low photosensitivity, short lifetime in the electrophotographic process, and concerns about the safety of TNF. It was eventually replaced with dual-layer OPCs with CGL materials based on pigments. Acceptor-doped PVK continues to receive attention particularly with C_{60} and related molecules as the acceptor [39.116].

The dye-polymer-aggregate based CGL is unique in that the active material is formed in situ during the coating process. A solution of thiapyrylium dye (2,6-diphenyl-4(4-dimethylaminophenyl)thiapyrylium perchlorate) and bisphenol-A polycarbonate is coated to form a CGL in which the dye and polymer spontaneously aggregate [39.117–119]. The aggregate is a dense filamentary structure which has an absorption spectrum red-shifted from that of the dye. The aggregate filaments

transport electrons with high field trapping due to carrier immobilization at filament dead ends. The photosensitivity of the aggregate is several orders of magnitude larger than that of the nonaggregated dye [39.120]. To prepare a CGL the formulation also includes hole-transporting materials. Because the CGL transports both holes and electrons it is possible to use the aggregate as a single layer or as the CGL in a dual-layer photoreceptor. Because the aggregate absorption drops rapidly at wavelengths longer than 700 nm this material is only useful for visible-light exposure systems, Fig. 39.13. In this example the CGL (2 μm) comprised dye and tri-*p*-tolylamine (TTA) hole-transport material in polycarbonate, and the CTL (13 μm) was a solid solution of 40 wt % tri-*p*-tolylamine in a polyester binder [39.88]. This was charged to -600 V and exposed with very low-intensity continuous irradiation to obtain the initial rate of photodischarge. At 680 nm the photogeneration efficiency is 0.34 per absorbed photon. The quantum efficiency (Onsager model) was 0.6 and the thermalization distance was dependent on the concentration of hole-transport material ranging from 200 nm (5 wt %) to 600 nm (40 wt %) [39.88], Fig. 39.11.

Azo compounds contain the $-\text{N}=\text{N}-$ functionality. Bisazo compounds have the general structure $\text{R}'-\text{N}=\text{N}-\text{R}-\text{N}=\text{N}-\text{R}''$. Trisazo compounds have the general structure $\text{R}(-\text{N}=\text{N}-\text{R}')(-\text{N}=\text{N}-\text{R}'')(-\text{N}=\text{N}-\text{R}''')$. The R substituents generally have extended conjugation to provide the colors which have made bisazo compounds important dyes and pigments for many years. Bisazo CGLs generally absorb in the visible and trisazo into the near-infrared. An early example of a bisazo pigment-based OPC (chlorodiane blue, 4,4'-bis(1''-azo-2''-hydroxy-3''-naphthylidene)-3,3'-dichlorobiphenyl) was used in the IBM 3800 printer. This photoreceptor had over four times the photosensitivity of the PVK-TNF-based OPC [39.121]. Azo charge-generation materials are often synthesized from the reaction of an aromatic amine with an o-hydroxyaromatic carboxylic acid such as naphthol-AS. As CGL materials these materials are used as dispersed pigments. Although bisazo pigments do not generally exhibit polymorphism they can exist as hydroxy-azo or keto-hydrazone tautomers. In some cases, it has been found that the active form is the keto-hydrazone tautomer [39.122]. Because the synthesis of azo compounds is relatively easy there is a considerable literature on the effects of chemical structure on electrophotographic characteristics and carrier-generation efficiency [39.123–127]. In azo pigments charge is generated extrinsically with the pigment

exciton oxidizing a hole-transport molecule at the pigment surface [39.127, 128] and it was demonstrated that penetration of the CTL into the CGL occurs during the coating process [39.129]. The photosensitivity action spectrum for a fluorenone bisazo pigment is shown in Fig. 39.13 [39.128]. In this OPC the CGL ($0.17\text{ }\mu\text{m}$) consisted of the pigment in a binder of poly(vinyl butyral) (10 : 4 weight ratio) and the CTL ($17.4\text{ }\mu\text{m}$) a solid solution of hydrazone transport material, *p*-diethylaminobenzaldehyde diphenylhydrazone (DEH), in polycarbonate (9 : 10 weight ratio). The initial potential was -800 V and the exposure required for a discharge to -400 V determined. The quantum yield of carrier generation from the Onsager model was 0.52 with a thermalization length of 1400 nm . Time-resolved transient absorption spectroscopy has been used to study the formation and decay of the oxidized transport material [39.130]. With a bisazo-based CGL (chlorodiane blue) [39.131] and trisazo-based CGL it was found that the charge-generation efficiency depended on the oxidation potential of the CTL transport materials [39.125]. In the most favorable case the bisazo had a quantum efficiency of 0.25 and the trisazo about 0.5. This was interpreted in terms of the energy gap between the hole at the pigment surface and the transport material. In a study using a series of bisazo pigment CGLs and a common CTL it was found that neither electron donating nor withdrawing substituents favor high photosensitivity [39.126]. Photosensitivity was found to depend inversely on the crystallinity of the pigment, which was interpreted in terms of pigment surface area. The development of high-sensitivity azo-based CGLs has been reviewed by Murayama [39.132]. The trisazo pigment-based OPC has an absorption which extends into the near-infrared, making it suitable for laser exposures. Figure 39.14 shows the photosensitivity action spectrum of an OPC with a trisazo pigment CGL [39.133]. The CGL ($0.33\text{ }\mu\text{m}$) consisted of the pigment in poly(vinyl butyral) binder (10 : 4 weight ratio) and the CTL ($18.5\text{ }\mu\text{m}$) was a solid solution of DEH transport material in polycarbonate (9 : 10 weight ratio). The initial surface potential was -700 V and the exposure to discharge to -300 V used to determine the photosensitivity at each wavelength. A photoreceptor with CGL ($0.12\text{ }\mu\text{m}$) and CTL ($19.5\text{ }\mu\text{m}$) had a quantum efficiency for charge generation of 0.46 from $500\text{--}800\text{ nm}$ [39.133].

Perylenes used as CGL materials are diimides of perylene-3,4,9,10-tetracarboxylic acid. One example is the *N,N'*-bis(2-phenethyl)-perylene-3,4:9,10-bis(dicarboximide). In this case the CGL was prepared

by vacuum deposition of an amorphous film ($0.1\text{ }\mu\text{m}$). Exposure to dichloromethane (as when overcoating with the charge transport layer) converts the film into a crystalline form with enhanced absorption at 620 nm [39.135]. The mechanism of carrier generation of this pigment in the presence of tri-*p*-tolylamine has been studied by studying the field dependence of quenching of perylene fluorescence [39.136]. The results were interpreted in terms of both intrinsic (directly from the perylene singlet state) and sensitized (interaction of holes on the pigment surface with the amine) charge generation. More recently, a bisbenzimidazole perylene has been developed as a CGL [39.137–139]. Dual-layer OPCs with a CGL of $0.1\text{-}\mu\text{m}$ vacuum-deposited bisbenzimidazole perylene, and a $15\text{-}\mu\text{m}$ CTL of 35 wt % TPD in polycarbonate, exhibit high sensitivity ($E_{1/2}$ 2 erg/cm^2 for -800 V_0) out to 700 nm , Fig. 39.13. Recent research has involved the synthesis and study of novel materials such as dimeric perylene pigments [39.140].

Phthalocyanines have received considerable attention as CGL materials for near-infrared exposure systems. Metal-free (H_2Pc) [39.141–144], titanyl (TiOPc) [39.134], and hydroxygallium (HOGaPc) [39.145] phthalocyanines have been extensively stud-

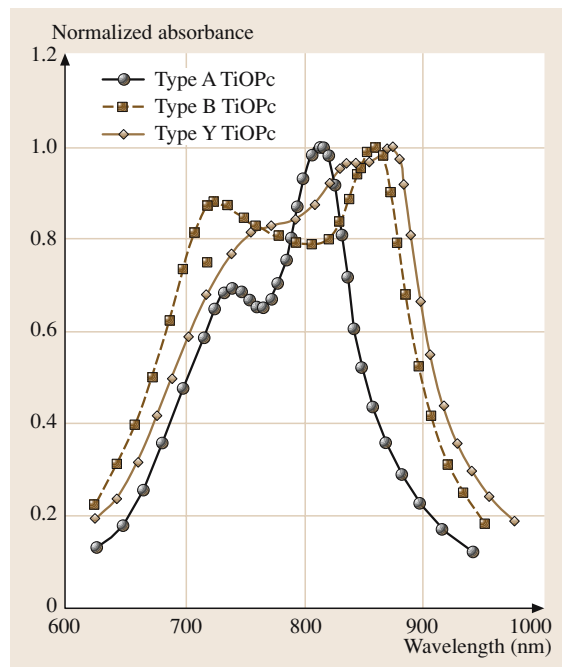


Fig. 39.17 Absorption spectra of polymorphs of TiOPc. After [39.134]

ied as CGL pigments. Although these materials are relatively easy to synthesize their intractability makes them difficult to purify. Except for halogen substituents on the aromatic rings there has been little research into the effect of chemical structure on performance. Instead, much of the development has been to understand and control polymorphism. Compared with solution absorption spectra, the spectra of many CGL pigments are broadened with enhanced absorption in the red and near-infrared. With specific combinations of milling technique, solvent exposure, and thermal treatment many phthalocyanine polymorphs can be produced [39.146–148]. Polymorphs have been characterized by visible and near-infrared absorption spectroscopy, X-ray diffraction, nuclear magnetic resonance (NMR) spectroscopy, infrared spectroscopy, and infra-red (IR) absorption spectroscopy. Examples of TiOPc polymorph absorption spectra are shown in Fig. 39.17 [39.134]. Polymorphs may have very different photogeneration characteristics as CGLs in OPCs. The materials shown in Fig. 39.17 were fabricated into OPCs: CGL, 0.4 μm , pigment and poly(vinyl butyral) 1 : 1 weight ratio; CTL, 20 μm , 43 wt % triarylamine hole-transport material, 4-(4-methylstyryl)-4',4''-dimethoxytriphenylamine, in polycarbonate. The photosensitivity action spectra for OPCs from the Y- and A-forms of TiOPc are shown in Fig. 39.14. The OPC with the Y-form had a photosensitivity for half photodischarge (600–850 nm) of 0.75 erg/cm² (initial potential –600 V) while the A-form was much less sensitive. In another study, photoreceptors formulated with α -, β -, τ -, and X-forms of H₂Pc had generation efficiencies as follows (5×10^5 V/cm): $\tau_2 \approx 0.1$, $X_2 \approx 0.09$, $X_1 \approx 0.03$, $\beta_1 \approx 0.01$, $\alpha_2 \approx 0.005$, and $\alpha_1 \approx 0.001$ [39.149]. In the preparation of a generation material a common procedure is to prepare the pigment in the amorphous form by acid pasting or with a specific milling technique, and then to convert the pigment to the desired polymorph with a specific solvent treatment. The desired polymorph must of course be stable to all the subsequent operations. In addition to the dominant influence of crystal form on electrophotographic performance, recent studies have implicated water in carrier generation of Y-form TiOPc [39.134, 150]. A recent paper has reviewed the carrier generation mechanisms in several phthalocyanines [39.151]. In all the pigments studied [x -H₂Pc, TiOPc (I) (low sensitivity), TiOPc (IV) (also known as Y, high sensitivity), HOGaPc (V)], fluorescence quenching implicates a neutral excited state (not a charge-transfer state) as the precursor to carrier generation. Studies of

the effect of field on fluorescence amplitude and lifetime reveal that with HOGaPc and TiOPc (I and IV) carrier generation comes from a non-relaxed excited state and TiOPc (IV) has an additional generation channel from a trapped state. With x -H₂Pc trapped excitons appear to be the major source of charge carriers.

Squarylium (squaraine) pigments are a class of materials which were first synthesized only recently. Squaryliums [2,4-di-(*o*-hydroxy-*p*-dimethylaminophenyl)cyclobutene and 1,4-bis-(4-dimethylamino-2-methylphenyl)diylum-1,3-diolate] were the first pigments to be utilized as a near-infrared-sensitive CGL in an IBM printer [39.131, 152, 153]. The photoreceptor based on the hydroxy squarylium with a CTL (21 μm , DEH in bisphenol-A polycarbonate) had high photosensitivity from 500–800 nm (≈ 3 erg/cm² for a –700 to –200 V photodischarge), Fig. 39.14 [39.153]. This class of pigment has been extensively investigated for use as pigment-dispersion CGLs for near-infrared exposure systems [39.137, 154–160]. A nonsymmetrical squaraine (UFSq)-based OPC with CGL, 0.4- μm UFSq in poly(vinyl formal) (80 : 20 weight ratio), and CTL 26- μm 40 wt % TPD in bisphenol-A polycarbonate, exhibited an $E_{1/2}$ photosensitivity of 1.9 erg/cm² at 790 nm ($-980 V_0$) when optimally purified [39.160], Fig. 39.14.

39.4.6 Charge-Transport Layer (CTL)

In all current commercial organic photoreceptors holes are the carriers that dominate photodischarge of the device. Hole transport is carried out by aromatic amines which are either doped into, or are incorporated as part of, the binder polymer. The key molecular characteristic is that aromatic amines have relatively low oxidation potentials. The oxidation potential of the transport moiety must be lower than that of the charge-generation material such that photogenerated holes are injected from the generation material into the charge-transport layer. In a chemical sense the key reaction is oxidation of the charge-transport material by the photogenerated hole on the surface of the generation material. If the oxidation potential of the transport material is too low it will undergo air oxidation, if it is too high its oxidation will require an activation energy and will be inefficient. Molecular orbital studies suggest that molecules with little change in geometry with oxidation will be favorable for hole transport [39.161]. Aromatic amines with oxidation potentials between about 0.8–1.2 eV (relative to SCE) are generally useful as hole-transport materials. The synthesis and study of hole-transport molecules

is a very active area. Figure 39.4 shows the chemical structures of some common hole-transport materials: 1,1-bis(*p*-diethylaminophenyl)-4,4-diphenyl-1,3-butadiene (DPB), *p*-diethylaminobenzaldehyde diphenylhydrazone (DEH), tri-*p*-tolylamine (TTA), N,N'-diphenyl-N,N'-bis(3-methylphenyl)-(1,1'-biphenyl)-4,4'-diamine (TPD), and 1,1-bis(di-4-tolylaminophenyl)cyclohexane (TAPC). These are often referred to by acronym and classified according to one of their structural functionalities: arylmethane, diarylamine, enamine, triarylamine, heterocyclic, butadiene, and hydrazone. A more complete listing is available in *Borsenberger and Weiss* [39.4].

Charge transport was discussed previously (Sect. 39.2.2). Since transport involves orbital overlap between neighboring charge-transport moieties their concentration needs to be as high as possible. In practice, molecular transport materials are doped, singly or as a mixture, at 40–50% by weight into the binder polymer. The purpose of the binder polymer is to provide an inert, robust, substrate for the transport material. The mobility characteristics of a large number of hole-transport materials have been studied and the results interpreted according to a model where the dissolved transport molecules are disordered in energy and position [39.162–164]. Binder effects on hole transport have been observed. It appears that the presence in the CTL of materials with highly polar substituents decrease hole mobility but in general these effects are not well understood [39.165–170].

Since the CTL is typically the surface that is exposed to the various electrophotographic process elements it must be formulated to resist damage and wear. The glass-transition temperature (T_g) of the CTL is an important characteristic. Too low and the material will be soft and easily damaged, too high and it will be brittle. Generally a T_g around 60 °C is adequate. Polycarbonates and polyesters are two classes of materials commonly used as CTL binder polymers.

Dual-layer devices in which electron transport dominates have been studied extensively but never commercialized. Examples of commercialized electron-transport-dominated photoreceptors are the single-layer dye-polymer aggregate and the PVK-TNF charge-transfer complex. Electron-transport materials are designed to be easily reduced. The reduction potential must be lower than that of the photogenerated electron on the charge-generation material so that it is transferred to the transport material. However, it is believed that oxygen may act as an electron trap so the reduction potential of the transport material must be lower than

that required to reduce oxygen dissolved in the polymer film. Figure 39.10 shows the chemical structures of some electron-transport materials. As with hole-transport materials acronyms based on molecular functionality are common: 1,1-dioxo-2-(4-methylphenyl)-6-phenyl-4-(dicyanomethylidene)thiopyran (PTS), *n*-butyl 9-dicyanomethylenefluorene-4-carboxylate (BCMF), 2,4,7-trinitro-9-fluorenone (TNF), 3,3'-dimethyl-5,5'-di-*t*-butyl-4,4'-diphenoquinone and 3,5'-dimethyl-3',5-di-*t*-butyl-4,4'-diphenoquinone isomer mixture (DPQ); a more complete table can be found in *Borsenberger and Weiss* [39.4]. In general the charge mobilities observed with electron-transport materials are much lower than for hole-transport materials. For this reason dual-layer OPCs with electron-transport-based CTLs have not been commercialized. Analysis of the mobility characteristics of electron transport with the disorder model implicates the highly polar nature of all electron-transport molecules [39.171–174].

Both hole- and electron-transport materials have been prepared with the transport moieties as part of a polymer (main chain or pendant) which would serve as a binder with hole-transport functionality. The advantage of a transport-active polymer is that it is a single material. In practice, charge transport is generally not enhanced in these materials and their chemical complexity makes them expensive to synthesize. Other than PVK, molecularly doped polymers are the current choice for CTLs.

39.4.7 Backing Layer

In a web photoreceptor the substrate-photoreceptor may tend to curl due to the residual stress of the coated layers. Such curl can make it difficult to fabricate the web into loops, may degrade loop tracking in the machine, or cause the loop edge to touch other process elements. Any of these attributes would make the photoreceptor unusable. One way to counter film curl is to coat a polymer layer on the back side to compensate for the stresses in the photoreceptor layers. The back-coat polymer must have adequate physical characteristics (friction, wear, etc.) to withstand cycling in the process. The extra coating adds expense but its addition is relatively commonplace. Any suitable polymer, such as a polycarbonate, may be used.

39.4.8 Overcoat Layer

Overcoat layers have received increased attention to lengthen the photoreceptor life by preventing wear, min-

imizing the effects of surface contacts, the effects of corona effluents (ozone and nitrogen oxides), and liquid attack in liquid-toner-based printers. As with the charge-blocking layer, the simplest approach is to coat a thin insulating polymer onto the photoreceptor surface. The thickness cannot be more than a few μm if one wants to avoid large residual potentials. Several approaches have been developed. One approach is to coat a second CTL with a changed composition to maximize polymer toughness (often by lowering the concentration of transport material). Other approaches are to overcoat the photoreceptor with a silsesquioxane sol-gel several μm thick, or with a submicron refractory layer of diamond-like carbon or silicon nitride.

Silsesquioxane overcoats applied in a sol-gel process have received recent attention [39.175, 176]. One commercial application utilized a sol-gel to overcoat a photoconductor with an uppermost CGL (positive-charging OPC). Because the CGL is sensitive to wear

and corona chemicals an overcoat was essential to its implementation in a commercial photoreceptor. The sol is prepared and subsequently hopper-coated and cured at elevated temperature to effect crosslinking. To avoid residual potential some conductivity is necessary but this cannot be so high that the latent image is degraded [39.177].

A polyurethane overcoat in a positive-charging dual-layer OPC has recently been described for use in a liquid-toner-based printer [39.178]. The CGL was based on TiOPc pigment and, to avoid disruption of the CGL during the overcoat coating operation, an aqueous polyurethane dispersion was used.

Diamond-like carbon, silicon nitride, and other [39.179] overcoats are applied in a vacuum process as submicron layers. Although tough, their thinness limits their utility for scratch and wear resistance. Furthermore, depending on their chemical make up they may have sensitivity to corona gasses.

39.5 Photoreceptor Fabrication

As discussed above the two most common photoreceptor configurations are as a loop or a metal drum. After coating the layers of a drum photoreceptor it is ready to use in the process. Generally the drums are inspected for visible defects and samples are taken for process testing and quality control. Since the substrate is conductive the ground contact is easily made by means of a conductive brush contacting the edge or inside of the drum when it is mounted in the machine. On the other hand, after the appropriate layers are coated on a web several more steps are required for fabrication of the photoreceptor. The web is rolled onto a spool at the end of the coating process and subsequently slit to the desired width and rewound. The coated roll is then inspected for visible defects and defective areas are marked for removal. Defective areas are removed when the photoreceptor is cut to the desired length. The cut sheets are then ultrasonically welded to produce a photoconductor loop. Because the conductive layer is buried between the substrate and the coated layers, making a ground contact is not trivial. It is possible to leave one edge of the ground layer uncoated but, because the thin metal would be readily damaged or worn away with use, one typically finds a conductive stripe coated on one edge of the photoreceptor. The only requirement is that the conductive stripe penetrate the coated layers and contact the metal. The ground contact in the machine is often

made with a conductive brush contacting the conductive stripe. The printing machine must keep track of the splice for several reasons: cleaning blade, degraded photoconductivity due to the ultrasonic welding, etc. This is sometimes accomplished by applying marks to the back of the photoreceptor, which are detected by the machine.

Fabricating a photoreceptor loop entails several operations not required in the fabrication of drum photoreceptors (slitting, cutting, conductive striping, welding, backside marking, etc.). Web-coating a photoreceptor is an expensive process but many thousands of feet of photoreceptor can be produced in a single coating event. Furthermore all the loop-processing elements are relatively straightforward and inexpensive. The key factor is that web-coating is a continuous process while drum-coating is a batch process. Photoreceptor loops are generally about 10 \times less costly to produce than drums.

There are advantages and disadvantages to both seamed loop and drum configurations in a printing machine. Loops can be made small for desktop printers or large for high-speed commercial printers. A large web (current commercial webs are as large as ≈ 10 feet in circumference) provides more surface area for imaging (several images per loop cycle) and more space for the various electrophotographic process elements. In addition, exposures through a semitransparent conductive layer are possible. Disadvantages are the presence of

a seam which must be tracked and makes some processes, such as cleaning with a blade, problematic. Also, the tendency of a web to flutter or move necessitates both tracking and positioning to achieve adequate print registration. Metal drums are seamless and rigid, making it easier to control drum position. However, drums

typically have a diameter of 35–300 mm and have limited imaging area and space for process elements. All things being equal a loop which produces several images per cycle will produce more images over its lifetime than a drum. Ultimately, the choice of loop or drum is a printing system issue.

39.6 Summary

Organic photoreceptors have been utilized in electrophotography since 1970. In this chapter we have briefly reviewed the development of organic photoreceptors from the standpoint of the device architecture, materials formulation, and electrical characteristics with emphasis on how the materials used impact on specific device functions.

The use of organic materials has enabled the development of electrophotographic technologies into a multi-billion-dollar business. Today's organic photoreceptor is a large-area photoelectric device capable of very high spatial resolution in the formation of a large-area image, with pixel resolution as small as 20 μm (1200 dpi), to meet demanding image-quality requirements. It is truly an amazing success story. Many of the principles and indeed many of the same materials are being applied to today's highly publicized new technologies including organic light-emitting diodes (OLED) and organic transistors. In this chapter we have described the device requirements and shown how organic photoreceptor development has progressed through a combination of architecture manipulation and materials design.

Organic photoreceptors have to satisfy many diverse requirements to be successfully employed in electrophotography as it is practiced today. These requirements have been met in large part by the principles of: (1) functional separation, and (2) materials design for the enhancement of specific electrical and mechanical characteristics. The concept of functional separation was key to organic photoreceptor development. Rather than try to produce a device with all the desirable characteristics in a single layer it is more productive to have a multi-layer architecture with each layer being optimized for a specific characteristic. Not surprisingly these layers are commonly named after those characteristics which are paramount: the charge-generation layer (CGL), the charge-transport layer (CTL), the charge-blocking layer, the protective overcoat layer, etc. In this chapter we have described the characteristics of each of these layers and

demonstrated how materials have been chosen and new materials synthesized to optimize their characteristics.

It is worthwhile to take a minute to review the electrophotographic process and consider the characteristics which have been built into the device.

- **Spatial uniformity.** The thickness of each of the coated layers (each varies from submicron to a few tens of μm) and the total thickness (controlled mainly by the CTL) must be very uniform so that the device capacitance (a major factor in determining surface charging characteristics) is uniform (on the order of a few percent variability). The CGL generation characteristics must be uniform, which means that the optical density of the CGL at the exposure wavelength must be uniform, to produce a uniform surface potential after exposure (on the order of a few percent variability). Thickness must be uniform for the smallest pixel (20 μm for 1200-dpi imaging) over the entire area of the device ($> 1000 \text{ cm}^2$). The ability to coat large areas of such thin layers to this degree of uniformity is a great technological achievement.
- **Chemical resistance.** Photoreceptors must resist chemical attack. Because corona charging is the most common method of applying a surface potential, the photoreceptor must be stable to the acid and oxidizing species produced in the corona. These chemicals (ozone, NO_x , and HNO_x) are highly reactive and the layer that is exposed (usually the CTL) must not be degraded. This is achieved by judicious choice of materials, the addition of stabilizers (antioxidants) to the CTL, or a protective overcoat.
- **Electrical uniformity and stability.** The photoreceptor dark electrical and photoelectrical characteristics (surface potentials with no exposure, imaging exposure, and erase exposure) must be very uniform and stable to prolonged electrophotographic cycling. The initial (dark) and exposed potentials must be the same for each pixel over the entire device area and

must remain stable with cycling. If these characteristics drift uniformly with cycling, machine process control must be implemented to maintain image quality. If these characteristics drift nonuniformly the photoreceptor must be discarded when image artifacts (ghosting) appear.

- **Photosensitivity.** The sensitivity of the photoreceptor to light, often expressed as the energy needed to produce a defined decrease in surface potential, must be mated with the exposure used in the process in terms of exposure energy and wavelength. This has been achieved by the synthesis and pigment form manipulation of charge-generation materials. Today's high-sensitivity photoreceptors can approach a quantum efficiency of unity for charge generation (at a typical applied field of ≈ 20 V/cm). But, most importantly photoreceptors with lower sensitivity can be fabricated depending on the needs of the imaging system.
- **Photodischarge rate (switching time).** The time it takes for photogenerated carriers to transit the device limits the printing process speed. With the synthesis of highly stable hole-transport materials and the principle of molecularly doped polymers it has been possible to prepare devices with hole transit times that permit printing speeds approaching 200 pages/min.
- **Mechanical strength.** The surface of the photoreceptor that is exposed to the abrasive process elements

(often the development and cleaning systems), must not wear too rapidly or unevenly and must be stable to crazing or other forms of damage. Typically the CTL is the surface layer, and considering that it is typically a polymer with as much as 50% by weight of charge-transport molecules, it is remarkable that materials can be chosen such that significant wear occurs only after tens of thousands of cycles. Where warranted, very tough protective overcoats have been developed which have little effect on the overall OPC device performance.

Today multilayer organic photoreceptors that meet all of the requirements listed above are made commercially on a huge scale. This success is due in large measure to the skill of the scientists and engineers who design the materials and processes for photoreceptor fabrication, and who then integrate it into the electrophotographic printing system such that acceptable and consistent image quality is obtained for a photoreceptor life, which is routinely 200 000 imaging cycles or longer if protective overcoat layers are used. This is an amazing story, an organic photoreceptor device with *adjustable* sensitivity at any desired wavelength from the visible to the near-infrared, with stable dark and photoelectrical characteristics over many thousands of imaging cycles, such that the characteristics of a 40- μ m pixel area are identical over an area which might be greater than 1000 cm².

References

- 39.1 J. Mort, D. M. Pai: *Photoconductivity and Related Phenomena* (Elsevier, New York 1976)
- 39.2 N. V. Joshi: *Photoconductivity Art, Science, and Technology* (Marcel Dekker, New York 1990)
- 39.3 K.-Y. Law: *Chem. Rev.* **93**, 449 (1993)
- 39.4 P. M. Borsenberger, D. S. Weiss: *Organic Photoreceptors for Xerography* (Marcel Dekker, New York 1998)
- 39.5 P. M. Borsenberger, D. S. Weiss: Photoreceptors: Organic Photoreceptors. In: *Handbook of Imaging Materials*, ed. by A. S. Diamond, D. S. Weiss (Marcel Dekker, New York 2002) p. 369
- 39.6 D. S. Weiss, J. R. Cowdery, R. H. Young: Electrophotography. In: *Molecular-Level Electronics, Imaging and Information, Energy and Environment*, 2, Vol. 5, ed. by V. Balzani (Wiley-VCH, Weinheim 2001) Chap. 2
- 39.7 R. M. Schaffert, C. D. Oughton: *J. Opt. Soc. Am.* **38**, 991 (1948)
- 39.8 J. H. Dessauer, G. R. Mott, H. Bogdonoff: *Photogr. Eng.* **6**, 250 (1955)
- 39.9 J. Mort, I. Chen: *Appl. Solid State Sci.* **5**, 69 (1975)
- 39.10 J. Mort: *The Anatomy of Xerography* (McFarland, Jefferson 1989)
- 39.11 C. F. Carlson: *Electrophotography*, US Patent 2 297 691 (1942)
- 39.12 I. Chen: *J. Imaging Sci.* **34**, 15 (1990)
- 39.13 I. Chen: Nature of Latent Images Formed on Single Layer Organic Photoreceptors. In: *Proc. ISE/T NIP18: 2002 Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2002) p. 404
- 39.14 S. Jeyadev, D. M. Pai: *J. Imaging Sci. Technol.* **40**, 327 (1996)
- 39.15 D. M. Pai, B. E. Springett: *Rev. Mod. Phys.* **65**, 163 (1993)
- 39.16 L. B. Schein: *Electrophotography and Development Physics* (Laplacian, Morgan Hill 1996)

- 39.17 R. M. Schaffert: *Electrophotography* (Focal, New York 1975)
- 39.18 M. E. Scharfe: *Electrophotography Principles and Optimization* (Wiley, New York 1984)
- 39.19 E. M. Williams: *The Physics and Technology of Xerographic Processes* (Wiley-Interscience, New York 1984)
- 39.20 R. G. Enck, G. Pfister: Amorphous Chalcogenides. In: *Photoconductivity and Related Phenomena*, ed. by J. Mort, D. M. Pai (Elsevier, New York 1976) p. 215
- 39.21 S. O. Kasap: Photoreceptors: The Chalcogenides. In: *Handbook of Imaging Materials*, ed. by A. S. Diamond, D. S. Weiss (Marcel Dekker, New York 2002) p. 329
- 39.22 J. Mort: Applications of Amorphous Silicon and Related Materials in Electronic Imaging. In: *Handbook of Imaging Materials*, ed. by A. S. Diamond, D. S. Weiss (Marcel Dekker, New York 2002) p. 629
- 39.23 R. Joslyn: Photoreceptors: Recent Imaging Applications for Amorphous Silicon. In: *Handbook of Imaging Materials*, ed. by A. S. Diamond, D. S. Weiss (Marcel Dekker, New York 2002) p. 425
- 39.24 A. Rose: *Photoconductivity and Related Processes* (Interscience, New York 1963)
- 39.25 J. C. Scott, G. S. Lo: Dark Decay in Organic Photoconductors. In: *Proc. 6th Int. Symp. Adv. Non-Impact Printing Technol.*, ed. by R. J. Nash (Society Imaging Science and Technology, Springfield 1991) p. 403
- 39.26 R. H. Bube: *Photoelectronic Properties of Semiconductors* (Cambridge Univ. Press, Cambridge 1992)
- 39.27 S. M. Sze: *Physics of Semiconductor Devices*, 2 edn. (Wiley, New York 1981)
- 39.28 M. A. Abkowitz, M. Stolka: Electronic Transport in Polymeric Photoreceptors: PVK in Polysilylenes. In: *Proc. Int. Symp. Polym. Adv. Technol.*, ed. by M. Lewin (VCH, Weinheim 1988) p. 225
- 39.29 D. M. Pai, J. F. Yanus, M. Stolka: *J. Phys. Chem.* **88**, 4714 (1984)
- 39.30 M. Stolka, J. F. Yanus, D. M. Pai: *J. Phys. Chem.* **88**, 4707 (1984)
- 39.31 M. Stolka, M. A. Abkowitz: *Mater. Res. Soc. Symp.* **277**, 3 (1992)
- 39.32 F. K. Dolezalek: Experimental Techniques. In: *Photoconductivity and Related Phenomena*, ed. by J. Mort, D. M. Pai (Elsevier, New York 1976) p. 27
- 39.33 A. R. Melnyk, D. M. Pai: Photoconductivity Measurements. In: *Physical Methods in Chemistry*, ed. by B. W. Rossiter, J. F. Hamilton, R. C. Baetzold (Wiley, New York 1992) p. 321
- 39.34 I. Chen: *J. Appl. Phys.* **43**, 1137 (1972)
- 39.35 I. Chen, J. Mort: *J. Appl. Phys.* **43**, 1164 (1972)
- 39.36 M. Abkowitz: *J. Reinforced Plastics Composites* **16**, 1303 (1997)
- 39.37 M. A. Abkowitz, M. Stolka, M. Morgan: *J. Appl. Phys.* **52**, 3453 (1981)
- 39.38 H. Bäessler: *Phys. Status Solidi (b)* **107**, 9 (1984)
- 39.39 M. Stolka, M. A. Abkowitz: TBD a Contribution to the Success of Organic Materials. In: *Practical Applications of Organic Electronic Materials*, ed. by M. Iwamoto, I. S. Pu, S. Taniguchi (Science Forum, Tokyo 1994) p. 265
- 39.40 C. W. Tang, S. A. VanSlyke: *Appl. Phys. Lett.* **51**, 913 (1987)
- 39.41 M. A. Abkowitz, H. Bäessler, M. Stolka: *Philos. Mag. B* **63**, 201 (1991)
- 39.42 W. D. Gill: *J. Appl. Phys.* **43**, 5033 (1972)
- 39.43 M. A. Abkowitz, J. S. Facci, W. W. Limburg, J. Janus: *Phys. Rev. B: Cond. Matter* **46**, 6705 (1992)
- 39.44 R. G. Kepler, J. M. Zeigler, L. A. Harrah, S. R. Kurtz: *Phys. Rev. B* **35**, 2818 (1987)
- 39.45 M. A. Abkowitz, M. Stolka: *Synth. Met.* **50**, 395 (1992)
- 39.46 D. Hertel, H. Baessler, U. Scherf, H. H. Horhold: *J. Chem. Phys.* **110**, 9214 (1999)
- 39.47 C. Im, H. Baessler, H. Rost, H. H. Horhold: *J. Chem. Phys.* **113**, 3802 (2000)
- 39.48 J. S. Facci, M. A. Abkowitz, W. W. Limburg, D. Renfer, J. Yanus: *Mol. Cryst. Liq. Cryst.* **194**, 55 (1991)
- 39.49 J. Michl, J. W. Downing, T. Karatsu, A. J. McKinley, G. Poggi, G. M. Wallraff, R. Sooriyakumaran: *Pure Appl. Chem.* **60**, 959 (1988)
- 39.50 R. D. Miller, J. Michl: *Chem. Rev.* **89**, 1359 (1989)
- 39.51 M. A. Abkowitz, M. Stolka: *Polym. Preprints (ACS Div. Polym. Chem.)* **31**, 254 (1990)
- 39.52 R. D. Miller, J. R. Rabolt, R. Sooriyakumaran, G. N. Fickes, B. L. Farmer, H. Kuzmany: *ACS Polym. Preprints* **28**, 422 (1987)
- 39.53 M. A. Abkowitz, M. Stolka: *Philos. Mag. Lett.* **58**, 239 (1988)
- 39.54 D. Emin: *Adv. Phys.* **24**, 305 (1975)
- 39.55 L. B. Schein, A. Rosenberg, S. L. Rice: *J. Appl. Phys.* **60**, 4287 (1986)
- 39.56 L. B. Schein: *Molec. Cryst. Liquid Cryst.* **183**, 41 (1990)
- 39.57 H. Bäessler: *Phys. Status Solidi* **175**, 15 (1993)
- 39.58 H. Baessler, G. Schoenherr, M. A. Abkowitz, D. M. Pai: *Phys. Rev. B: Cond. Matter* **26**, 3105 (1982)
- 39.59 P. E. Parris, V. M. Kenkre, D. H. Dunlap: *Phys. Rev. Lett.* **87**, 126601 (2001)
- 39.60 D. E. Dunlap, P. E. Parris, V. M. Kenkre: *Proc. SPIE* **3799**, 88 (1999)
- 39.61 L. B. Schein: *Philos. Mag. B* **65**, 795 (1992)
- 39.62 S. V. Novikov, A. V. Vannikov: *J. Phys. Chem.* **99**, 14573 (1995)
- 39.63 S. V. Novikov, D. H. Dunlap, V. M. Kenkre: *Proc. SPIE* **3471**, 181 (1998)
- 39.64 D. H. Dunlap: *Phys. Rev. B* **52**, 939 (1995)
- 39.65 D. H. Dunlap, P. E. Parris, V. M. Kenkre: *Phys. Rev. Lett.* **77**, 542 (1996)
- 39.66 S. V. Novikov, D. H. Dunlap, V. M. Kenkre, P. E. Parris, A. V. Vannikov: *Phys. Rev. Lett.* **81**, 4472 (1998)
- 39.67 M. A. Abkowitz: *Philos. Mag. B* **65**, 817 (1992)
- 39.68 M. Stolka, M. A. Abkowitz: *Synth. Met.* **54**, 417 (1993)
- 39.69 H. J. Yuh, D. Abramsohn, M. Stolka: *Philos. Mag. Lett.* **55**, 277 (1987)
- 39.70 F. W. Schmidlin: *Phys. Rev. B* **16**, 2362 (1977)

- 39.71 M. A. Abkowitz, R. C. Enck: J. Phys. Colloq. C **4**(1), 443 (1981)
- 39.72 M. A. Abkowitz: Characterization of Metal Interfaces to Molecular Media from Analysis of Transient and Steady State Electrical Measurements. In: *Conjugated Polymer and Molecular Interfaces*, ed. by W. R. Salaneck et al. (Marcel Dekker, New York 2002) p. 545
- 39.73 M. A. Abkowitz, J. S. Facci, M. Stolka: Chem. Phys. **177**, 783 (1993)
- 39.74 M. A. Abkowitz, J. S. Facci, M. Stolka: Appl. Phys. Lett. **63**, 1892 (1993)
- 39.75 M. A. Lampert, P. Mark: *Current Injection in Solids* (Academic, New York 1970) p. 17
- 39.76 P. N. Murgatroyd: J. Phys. D. **3**, 151 (1970)
- 39.77 R. H. Young: Philos. Mag. Lett. **70**, 331 (1994)
- 39.78 M. Pope, C. E. Swenberg: *Electronic Processes in Organic Crystals* (Clarendon, Oxford 1982)
- 39.79 Y. Hayashi, M. Kuroda, A. Inami: Bull. Chem. Soc. Jpn. **39**, 1660 (1966)
- 39.80 H. Hoegl: J. Phys. Chem. **69**, 755 (1965)
- 39.81 P. J. Regensburger: Photochem. Photobiol. **8**, 429 (1968)
- 39.82 H. Hoegl, H. G. Barchietto, D. Tar: Photochem. Photobiol. **16**, 335 (1972)
- 39.83 P. J. Melz: J. Chem. Phys. **57**, 1694 (1972)
- 39.84 M. Yokoyama, Y. Endo, H. Mikawa: Bull. Chem. Soc. Jpn. **49**, 1538 (1976)
- 39.85 M. Yokoyama, Y. Endo, A. Matsubara, H. Mikawa: J. Chem. Phys. **75**, 3006 (1981)
- 39.86 M. Yokoyama, S. Shimokihara, A. Matsubara, H. Mikawa: J. Chem. Phys. **76**, 724 (1982)
- 39.87 M. Yokoyama, H. Mikawa: Photogr. Sci. Eng. **26**, 143 (1982)
- 39.88 M. B. O'Regan, P. M. Borsenberger, E. H. Magin, T. Zubil: J. Imaging Sci. Technol. **40**, 1 (1996)
- 39.89 Z. D. Popovic: Carrier Generation Mechanisms in Organic Photoreceptors. In: *Proc. 9th Int. Cong. Adv. Non-Impact Printing Technol.* (Society Imaging Science and Technology, Springfield 1993) p. 591
- 39.90 L. Onsager: Phys. Rev. **54**, 554 (1938)
- 39.91 P. M. Borsenberger, A. I. Ateya: J. Appl. Phys. **50**, 909 (1979)
- 39.92 L. Onsager: J. Chem. Phys. **2**, 599 (1934)
- 39.93 C. Braun: J. Chem. Phys. **80**, 4157 (1984)
- 39.94 S. N. Smirnov, C. Braun: J. Imaging Sci. Technol. **43**, 425 (1999)
- 39.95 D. M. Pai: Geminate Recombination in some Amorphous Materials. In: *Physics of Disordered Materials*, ed. by D. Adler, H. Fritzsche, S. R. Ovshinsky (Plenum, New York 1985) p. 579
- 39.96 J. Noolandi, K. M. Hong: J. Chem. Phys. **70**, 3230 (1979)
- 39.97 T. Niimi, M. Umeda: J. Appl. Phys. **74**, 465 (1993)
- 39.98 Z. D. Popovic: J. Chem. Phys. **76**, 2714 (1982)
- 39.99 Z. D. Popovic: Chem. Phys. **86**, 311 (1984)
- 39.100 R. C. Hughes: J. Chem. Phys. **58**, 2212 (1973)
- 39.101 J. W. Kerr, G. H. S. Rokos: J. Phys. D: Appl. Phys. **10**, 1151 (1977)
- 39.102 I. Chen: J. Appl. Phys. **49**, 1162 (1978)
- 39.103 W. Mey, E. I. P. Walker, D. C. Hoesterey: J. Appl. Phys. **50**, 8090 (1979)
- 39.104 R. H. Young: J. Appl. Phys. **60**, 272 (1986)
- 39.105 J.-Y. Moisan, B. André, R. Lever: Chem. Phys. **153**, 305 (1991)
- 39.106 A. V. Buettner, W. Mey: Photogr. Sci. Eng. **26**, 80 (1982)
- 39.107 K. Aizawa, M. Takeshima, H. Kawakami: A Study of 1-dot Latent Image Potential. In: *Proc. NIP17: Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2001) p. 572
- 39.108 T. Iwamatsu, T. Toyoshima, A. Nobuyuki, Y. Muto, Y. Nakajima: A Study of High Resolution Latent Image Forming and Development. In: *Proc. ISE/T NIP 15: 1999 Internat. Conf. on Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 1999) p. 732
- 39.109 I. Chen: J. Imaging Sci. Technol. **37**, 396 (1993)
- 39.110 R. M. Schaffert: IBM J. Res. Devel. **15**, 75 (1971)
- 39.111 U. Vahtra, R. F. Wolter: IBM J. Res. Devel. **22**, 34 (1978)
- 39.112 J. M. Pearson: Pure Appl. Chem. **49**, 463 (1977)
- 39.113 R. C. Penwell, B. N. Ganguly, T. W. Smith: J. Polym. Sci., Makromol. Rev. **13**, 63 (1978)
- 39.114 W. D. Gill: Polymeric Photoconductors. In: *Photoconductivity and Related Phenomena*, ed. by J. Mort, D. M. Pai (Elsevier, New York 1976) p. 303
- 39.115 M. Hatano, K. Tanikawa: Prog. Organic Coat. **6**, 65 (1978)
- 39.116 F. Li, Y. Li, Z. Guo, Y. Mo, L. Fan, F. Bai, D. Zhu: Solid State Commun. **107**, 189 (1998)
- 39.117 W. J. Dulmage, W. A. Light, S. J. Marino, C. D. Salzberg, D. L. Smith, W. J. Staudenmayer: J. Appl. Phys. **49**, 5543 (1978)
- 39.118 P. M. Borsenberger, A. Chowdry, D. C. Hoesterey, W. Mey: J. Appl. Phys. **49**, 5555 (1978)
- 39.119 J. M. Perlstein: Structure and Charge Generation in Low-Dimensional Organic Molecular Self-Assemblies. In: *Electrical Properties of Polymers*, ed. by D. A. Seanor (Academic, New York 1982) p. 59
- 39.120 P. M. Borsenberger, D. C. Hoesterey: J. Appl. Phys. **51**, 4248 (1980)
- 39.121 D. McMurtry, M. Tinghitella, R. Svendsen: IBM J. Res. Devel. **28**, 257 (1984)
- 39.122 J. Pacansky, R. J. Waltman: J. Am. Chem. Soc. **114**, 5813 (1992)
- 39.123 G. DiPaola-Baranyi, C. K. Hsiao, A. M. Hor: J. Imaging Sci. **34**, 224 (1990)
- 39.124 K.-Y. Law, I. W. Tarnawskyj: J. Imaging Sci. Technol. **37**, 22 (1993)
- 39.125 M. Umeda, T. Shimada, T. Aruga, T. Niimi, M. Sasaki: J. Phys. Chem. **97**, 8531 (1993)
- 39.126 K.-Y. Law, I. W. Tarnawskyj, Z. D. Popovic: J. Imaging Sci. Technol. **38**, 118 (1994)
- 39.127 M. Umeda: J. Imaging Sci. Technol. **43**, 254 (1999)

- 39.128 M. Umeda, T. Niimi, M. Hashimoto: Jpn. J. Appl. Phys. **29**, 2746 (1990)
- 39.129 T. Niimi, U. Umeda: J. Appl. Phys. **76**, 1269 (1994)
- 39.130 K. Takeshita, Y. Sasaki, T. Shoda, T. Murayama: Time-Resolved Absorption Study on the Photocarrier Generation Process in Layered Organic Photoreceptors: A role of Delocalized Holes in Photocarrier Generation. In: *Proc. IS&T NIP19: 2003 Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2003) p. 683
- 39.131 P. J. Melz, R. B. Champ, L. S. Chang, C. Chiou, G. S. Keller, L. C. Licican, R. R. Neiman, M. D. Shattuck, W. J. Weiche: Photogr. Sci. Eng. **21**, 73 (1977)
- 39.132 T. Murayama: The Design of High Performance Organic Photoconductors. In: *Proc. IS&T NIP17: Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2001) p. 557
- 39.133 M. Umeda, M. Hashimoto: J. Appl. Phys. **72**, 117 (1992)
- 39.134 Y. Fujimaki, H. Tadokoro, Y. Oda, H. Yoshioka, T. Homma, H. Moriguchi, K. Watanabe, A. Konishita, N. Hirose, A. Itami, S. Ikeuchi: J. Imaging Technol. **17**, 202 (1991)
- 39.135 E. H. Magin, P. M. Borsenberger: J. Appl. Phys. **73**, 787 (1993)
- 39.136 Z. D. Popovic, R. Cowdery, I. M. Khan, A.-M. Hor, J. Goodman: J. Imaging Sci. Technol. **43**, 266 (1999)
- 39.137 R. O. Loutfy, A. M. Hor, P. Kazmaier: Properties and Application of Organic Photoconductive Materials: Molecular Design of Organic Photoconductive Polycyclic Aromatic Diimides Compounds. In: *Proc. 32nd Symp. Macromolecules*, ed. by T. Saegusa, T. Higashimura, A. Abe (Blackwell Scientific, Oxford 1988) p. 437
- 39.138 R. O. Loutfy, A. M. Hor, P. Kazmaier, M. Tam: J. Imaging Sci. **33**, 151 (1989)
- 39.139 Z. D. Popovic, A.-M. Hor, R. O. Loutfy: Chem. Phys. **127**, 451 (1988)
- 39.140 J. M. Duff, G. Allen, A.-M. Hor, S. Gardner: Synthesis, Spectroscopy and Photoconductivity of Dimetric Perylene Bisimide Pigments. In: *Proc. IS&T NIP15: 1999 Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 1999) p. 655
- 39.141 A. Kakuta, Y. Mori, S. Takano, M. Sawada, I. Shibuya: J. Imaging Sci. Technol. **11**, 7 (1985)
- 39.142 A. Shimada, M. Anzai, A. Kakuta, T. Kawanishi: IEEE Trans. Ind. Appl. **1A23**, 804 (1987)
- 39.143 Y. Kanemitsu, S. Imamura: J. Appl. Phys. **67**, 3728 (1990)
- 39.144 T. Enokida, R. Hirohashi, S. Mizukami: J. Imaging Sci. **35**, 235 (1991)
- 39.145 K. Daimon, K. Nukada, Y. Sakaguchi, R. Igarashi: J. Imaging Sci. Technol. **40**, 249 (1996)
- 39.146 T. Enokida, R. Hirohashi, T. Nakamura: J. Imaging Sci. **34**, 234 (1990)
- 39.147 S. Takano, Y. Mimura, N. Matsui, K. Utsugi, T. Gotoh, C. Tani, K. Tateishi, N. Ohde: J. Imaging Sci. Technol. **17**, 46 (1991)
- 39.148 T. I. Martin, J. D. Mayo, C. A. Jennings, S. Gardner, C. K. Hsiao: Solvent Induced Transformations of the Polymorphs of Oxytitanium Phthalocyanine (TiOPc). In: *Proc. IS&T Eleventh Int. Cong. Adv. Non-Impact Printing Technol.* (Society Imaging Science and Technology, Springfield 1995) p. 30
- 39.149 Y. Kanemitsu, A. Yamamoto, H. Funada, Y. Masumoto: J. Appl. Phys. **69**, 7333 (1991)
- 39.150 Z. D. Popovic, M. I. Khan, S. J. Atherton, A.-M. Hor, J. L. Goodman: Time-Resolved Fluorescence Quenching and Carrier Generation in Titanyl Phthalocyanine (TiOPc)-Humidity Effects. In: *Electrical and Related Properties of Organic Solids*, ed. by R. W. Munn, A. Miniewicz, B. Kuchta (Kluwer Academic, Dordrecht 1997) p. 207
- 39.151 Z. D. Popovic, M. I. Khan, A.-M. Hor, J. L. Goodman, J. F. Graham: Study of the Photoconductivity Mechanism in Phthalocyanine Pigment Particles by Electric Field Modulated Time Resolved Fluorescence. In: *Proc. IS&T NIP19: 2003 Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2003) p. 687
- 39.152 R. E. Wingard: IEEE Ind. Appl. **37**, 1251 (1982)
- 39.153 R. B. Champ: SPIE Proc. **759**, 40 (1987)
- 39.154 R. O. Loutfy, C. K. Hsiao, P. M. Kazmaier: Photogr. Sci. Eng. **27**, 5 (1983)
- 39.155 K.-Y. Law: J. Imaging Sci. **31**, 83 (1987)
- 39.156 K.-Y. Law, F. C. Bailey: J. Imaging Sci. **31**, 172 (1987)
- 39.157 R. O. Loutfy, A.-M. Hor, C.-K. Hsiao, G. Baranyi, P. Kazmaier: Pure Appl. Chem. **60**, 1047 (1988)
- 39.158 P. M. Kazmaier, R. Burt, G. DiPaola-Baranyi and C.-K. Hsiao, R. O. Loutfy, T. I. Martin, G. K. Hamer, T. L. Bluhm, M. G. Taylor: J. Imaging Sci. **32**, 1 (1988)
- 39.159 G. DiPaola-Baranyi, C. K. Hsiao, P. M. Kazmaier, R. Burt, R. O. Loutfy, T. I. Martin: J. Imaging Sci. **32**, 60 (1988)
- 39.160 K.-Y. Law: Chem. Mater. **4**, 605 (1992)
- 39.161 K. Sakanoue, M. Motoda, M. Sugimoto, S. Sakaki: J. Phys. Chem. **103**, 5551 (1999)
- 39.162 P. M. Borsenberger, E. H. Magin, M. Van der Auweraer, F. C. De Schryver: Phys. Stat. Sol. (a) **140**, 9 (1993)
- 39.163 P. M. Borsenberger, R. Richert, H. Bässler: Phys. Rev. B **47**, 4289 (1993)
- 39.164 M. Van der Auweraer, F. C. De Schryver, P. M. Borsenberger, H. Bässler: Adv. Mater. **6**, 199 (1994)
- 39.165 P. M. Borsenberger, H. Bässler: Phys. Status Solidi (b) **170**, 291 (1992)
- 39.166 A. Dieckmann, H. Bässler, P. M. Borsenberger: J. Chem. Phys. **99**, 8136 (1993)
- 39.167 P. M. Borsenberger, J. J. Fitzgerald: J. Phys. Chem. **97**, 4815 (1993)
- 39.168 R. H. Young: Philos. Mag. B **72**, 435 (1995)
- 39.169 A. Fujii, T. Shoda, S. Aramaki, T. Murayama: J. Imaging Sci. Technol. **43**, 430 (1999)

- 39.170 T. Shoda, T. Murayama, A. Fujii: Relationship Between Molecular Properties of Hole Transport Molecules and Field Dependence of Hole Mobility. In: *Proc. NIP17: Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2001) p. 550
- 39.171 P. M. Borsenberger, W. T. Gruenbaum: *J. Polym. Sci.: Polym. Phys.* **34**, 575 (1996)
- 39.172 P. M. Borsenberger, E. H. Magin, M. R. Detty: *J. Imaging Sci. Technol.* **39**, 12 (1994)
- 39.173 P. M. Borsenberger, H.-C. Kan, E. H. Magin, W. B. Vreeland: *J. Imaging Sci. Technol.* **39**, 6 (1995)
- 39.174 P. M. Borsenberger, W. T. Gruenbaum, M. B. O'Regan, L. J. Rossi: *J. Polym. Sci.: Polym. Phys.* **33**, 2143 (1995)
- 39.175 W. T. Ferrar, D. S. Weiss, J. R. Cowdery-Corvan, L. G. Parton: *J. Imaging Sci. Technol.* **44**, 429 (1999)
- 39.176 X. Jin, D. S. Weiss, L. J. Sorriero, W. T. Ferrar: *J. Imaging Sci. Technol.* **47**, 361 (2003)
- 39.177 D. S. Weiss, J. R. Cowdery, W. T. Ferrar, R. H. Young: *J. Imaging Sci. Technol.* **40**, 322 (1996)
- 39.178 N.-J. Lee, H.-R. Joo, K.-Y. Yon, Y. No: Development of Positive Charging Multi-Layered Organic Photoconductor for Liquid Electrophotographic Process. In: *Proc. ISE/T NIP19: 2003 Int. Conf. Digital Printing Technol.* (Society Imaging Science and Technology, Springfield 2003) p. 670
- 39.179 Y. C. Chan, X. S. Miao, E. Y. B. Pun: *J. Mater. Res.* **13**, 2042 (1998)

Luminescent

40. Luminescent Materials

This chapter surveys the field of solid-state luminescent materials, beginning with a discussion of the different ways in which luminescence can be excited. The internal energy-level structures of luminescent ions and centres, particularly rare-earth ions, are then discussed before the effects of the vibrating host lattice are included. Having set the theoretical framework in place, the chapter then proceeds to discuss the specific excitation process for photo-stimulated luminescence and thermally stimulated luminescence before concluding by surveying current applications, including plasma television screens, long-term persistent phosphors, X-ray storage phosphors, scintillators, and phosphors for white LEDs.

40.1	Luminescent Centres	985
40.1.1	Rare-Earth Ions	985
40.1.2	Transition-Metal Ions	986

40.1.3	s^2 Ions	987
40.1.4	Semiconductors	987
40.2	Interaction with the Lattice	987
40.3	Thermally Stimulated Luminescence	989
40.4	Optically (Photo-)Stimulated Luminescence	990
40.5	Experimental Techniques – Photoluminescence	991
40.6	Applications	992
40.6.1	White Light-Emitting Diodes (LEDs)	992
40.6.2	Long-Persistence Phosphors	992
40.6.3	X-Ray Storage Phosphors	993
40.6.4	Phosphors for Optical Displays	994
40.6.5	Scintillators	994
40.7	Representative Phosphors	995
	References	995

Luminescent materials are substances which convert an incident energy input into the emission of electromagnetic waves in the ultraviolet (UV), visible or infrared regions of the spectrum, over and above that due to

black-body emission. A wide range of energy sources can stimulate luminescence, and their diversity provides a convenient classification scheme for luminescence phenomena, which is summarised in Table 40.1. Pho-

Table 40.1 Types of luminescence

Designation	Excitation	Trigger	Acronym
Photoluminescence	UV, visible photons	–	PL
Radioluminescence	X-ray, gamma rays, charged particles	–	RL
Cathodoluminescence	Energetic electrons	–	CL
Electroluminescence	Electric field	–	EL
Thermoluminescence	Photons, charged particles	Heat	TSL
Optically/photo-stimulated luminescence	Photons, charged particles	Visible/IR photons	OSL, PSL

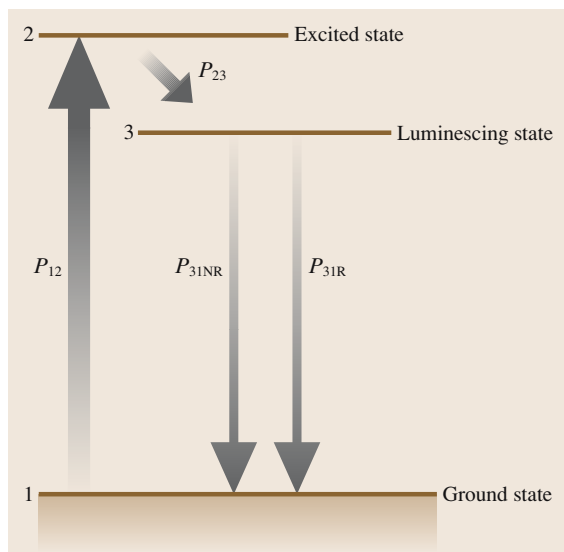


Fig. 40.1 Optical pumping cycle for a generic photoluminescent system

toluminescence, where the luminescence is stimulated by UV or visible light, is a widely used materials science technique for characterising dopants and impurities, and finds applications in lighting technologies such as fluorescent lamps. Radioluminescence involves excitation by ionising radiation, and is used in scintillators for nuclear particle detection; the special case of stimulation by energetic electrons is called cathodoluminescence, the name arising from early atomic physics experiments involving gas discharges. The major application of cathodoluminescence is in cathode ray tubes for television sets and computer monitors. Electroluminescence involves collisional excitation by internal electrons accelerated by an applied electric field, and with a much lower energy than in the case of cathodoluminescence. Electroluminescence finds applications in panel lighting used in liquid-crystal display (LCD) back-planes, and in light-emitting diodes.

There are other forms of luminescence which we mention for completeness but will not discuss further: bioluminescence and chemiluminescence where the energy input is from chemical or biochemical reactions, sonoluminescence (sound wave excitation), and triboluminescence (strain or fracture excitation).

There are several books which describe the luminescence of materials in more depth than is possible in a short article, for example that edited by Vij [40.1] and the monograph by Blasse and Grabmaier [40.2].

The terms phosphorescence and fluorescence are often used in connection with luminescent materials. This classification is based on the time-domain response of the luminescent system. Figure 40.1 shows a generic photoluminescent system where incident UV radiation excites a system from a ground state 1 with probability per unit time P_{12} into an excited state 2. The system decays with probability P_{23} to a luminescing level 3, from which there is a probability P_{31R} and P_{31NR} of radiative and nonradiative decay, respectively, to the ground state. Nonradiative decay generally involves phonon emission.

First suppose that the transition probabilities are such that $P_{12}, P_{21} \gg P_{31} (= P_{31R} + P_{31NR})$. If the optical pumping (excitation) is abruptly stopped, the population of the luminescing state 3 decays as,

$$N_3 = N_3(0) \exp(-P_{31}t) \quad (40.1)$$

and the rate of luminescent energy emission is $-(h\nu P_{31R}/P_{31})dN_3/dt$ for an energy difference $E_{31} = h\nu$ between states 1 and 3. Hence the luminescence intensity at distance r from the sample is,

$$I = (h\nu) N_3(0) P_{31R} \exp(-P_{31}t) / (4\pi r^2) \quad (40.2)$$

and the characteristic luminescence lifetime is $\tau = (P_{31})^{-1}$. Thus the lifetime is governed by both radiative and nonradiative processes, whilst the intensity of the luminescence depends on the relative magnitude of P_{31R} .

This discussion provides the basis for understanding the terms fluorescence and phosphorescence applied to luminescent materials. A material is often classified as one or the other according to the relative magnitude of $\tau = (P_{31})^{-1}$, with 10 ns being set in a relatively arbitrary way as the boundary between a *fast* fluorescent system and a *slow* phosphorescent one. For comparison, the theoretical lifetime for spontaneous emission for a strongly allowed hydrogen atom $2p \rightarrow 1s$ transition is about 0.2 ns.

However, by this definition phosphorescence can also arise from luminescent states with short lifetimes which are populated through ones with long lifetimes. In Fig. 40.1, if state 2 is long-lived in the sense that $P_{23} \ll P_{31}$ then the measured lifetime for emission from the luminescing state will be $\tau = (P_{23})^{-1}$, and the system will be labelled phosphorescent, even though the luminescing level itself has a very short lifetime. In consequence, a second classification [40.2] is based on whether or not the luminescing level is fed by a metastable state which sets the lifetime. Sometimes the metastable state is a long-lived intermediate form

of energy storage which can be triggered by an external stimulus to undergo a transition to a fluorescent level. Thus in Table 40.1, thermally stimulated and optically stimulated luminescence involve a metastable state consisting of trapped electrons and holes, which can be triggered to recombine by heating or by optical

stimulation; the recombination energy is transferred to a fluorescing centre.

Overall, the fluorescence/phosphorescence classification is somewhat nebulous, and it is debatable whether the classification is necessary or desirable.

40.1 Luminescent Centres

A wide variety of centres give rise to luminescence in semiconductors and insulating materials, including rare-earth ions, transition-metal ions, excitons, donor–acceptor pairs, and ions with a d^{10} or s^2 electronic configuration ground state. Some luminescence spectra consist of broad emission bands arising from the interaction between the electronic system of the luminescent centre and the vibrations of the atoms or ions, which surround it; the broad bands arise from simultaneous transitions of both electronic and vibrational systems. For others, such as the rare earths, the spectra comprise sharp lines arising from purely electronic transitions, and the effect of the environment is felt mainly through their effects on the lifetimes of the states. Thus in discussing the physical background to luminescence, it is simplest to start with a discussion of rare-earth luminescence, where the effect of vibrations can be initially ignored.

40.1.1 Rare-Earth Ions

The trivalent rare-earth ions have n electrons ($n = 1 - 14$) in the $4f$ shell. In a free ion, the eigenstates resulting from the various atomic interactions are labelled by the total spin S and orbital angular momenta L . Spin–orbit coupling breaks up each L, S multiplet of degeneracy $(2S + 1)(2L + 1)$ into sub-multiplets labelled by the total angular momentum $J = L + S$, where J can range from $L - S$ to $L + S$. The $4f^n$ orbitals lie within the outer $5s^2$ and $5p^6$ filled shells, which partly shield them from the effects of a crystalline environment. The effects of the latter are quantitatively described by the *crystal field* [40.3], and this term in the Hamiltonian splits the J multiplets into $2J + 1$ sublevels, the so-called crystal-field splitting. Some of these crystal-field levels may still be two- or threefold degenerate, depending upon the symmetry of the environment. Odd-electron systems always have at least twofold (Kramers) degeneracy. The resulting energy-level structures are complicated, and are summarised in the classic *Dieke di-*

agram [40.3]. The original has been updated by *Carnall et al.* [40.4] and is reproduced in many books and papers on rare-earth ion spectroscopy [40.2]. In Fig. 40.2, we show a schematic version (not accurately to scale) of the diagram appropriate for Pr^{3+} with $n = 2$, which serves for our discussion. The crystal-field splitting is usually smaller than the spin–orbit splitting, and is illustrated schematically by the vertical extent of the bands in Fig. 40.2. The multiplet labels follow the usual $^{2S+1}L_J$ system.

In the figure we also show a generic highest excited band which does not belong to the $4f^2$ configuration. In the specific case of the rare earths this can be the excited-state configuration resulting from promo-

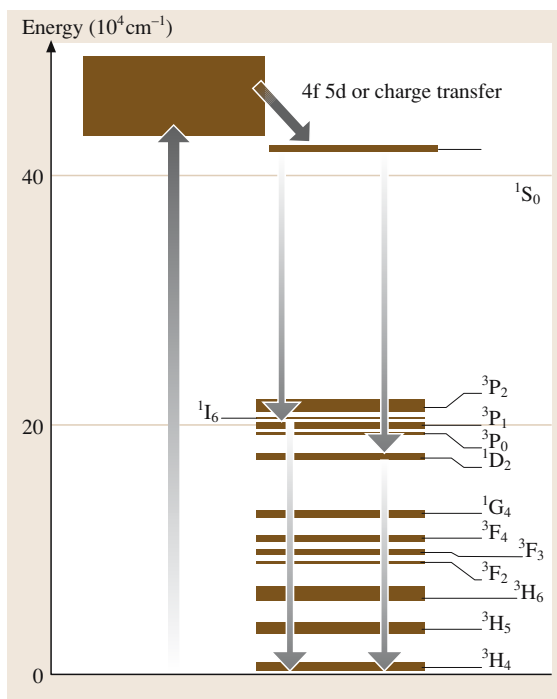


Fig. 40.2 Schematic energy-level diagram for Pr^{3+}

tion of one electron to the 5d state, giving an overall $4f^{n-1}5d^1$ configuration; more generally it could be a so-called charge transfer band, which corresponds to the transfer of one electron from the ligands to the luminescent ion. The relative location and importance of these bands varies with the luminescent ion and the crystalline environment, but they play an important role in the excitation of luminescence.

Excitation and luminescence transitions within the various levels in Fig. 40.2 are governed by the golden rule of quantum mechanics [40.5] for interactions with the electromagnetic field; in summary the probability of a transition between two states i and j is proportional to square of the matrix element $\langle i|H|j\rangle$, where H is the time-dependent perturbation Hamiltonian representing the interaction of the electrons with the electromagnetic field. The proportionality constant contains the light intensity in the case of excitation. The perturbation can be expanded in a power series involving electric and magnetic multipoles of the electronic system, but of these the electric dipole term is dominant, with the magnetic dipole term being much smaller by a factor of more than five orders of magnitude. Since the electric dipole operator \mathbf{er} has odd parity, the matrix element for transitions $r_{ij} = \langle i|\mathbf{er}|j\rangle$ is necessarily zero unless i and j have opposite parity. This is the most important selection rule governing luminescence: transitions between states of the same parity have zero transition probability and so are *forbidden* (Laporte's rule). In the case of the rare earths, all states of a single $4f^n$ configuration have the same parity, and so all optical transitions within the configuration are strictly forbidden. But this rule is relaxed by several considerations. First, if the crystalline environment lacks inversion symmetry, the crystal field admixes a small fraction of the excited configurations (eg $4f^15d^1$ for Pr^{3+}) of opposite parity into the ground configuration, which makes such transitions weakly allowed. Secondly, the selection rule for magnetic dipole transitions is that they are allowed between states of the same parity, although they are typically about five orders of magnitude weaker than for electric dipole ones. Finally, odd-parity vibrations and an electron-phonon interaction produce a similar configuration admixture effect to static lattice odd-parity mixing although this effect is more important for 3d ions than for 4f ones.

With only weak transitions possible within the 4f configuration, one might wonder how it would be possible to optically excite any significant luminescence. The answer lies in the 5d or charge-transfer bands which either lie at higher energies, or overlap with the upper levels of the 4f configuration. These give rise to strong

absorption and efficient pumping. Relaxation can occur via the parity-allowed transitions to the upper levels of the 4f configuration, and from there via single or multiple radiative emissions back to the ground state. Because these intra-configurational transitions are only weakly allowed, the lifetimes are generally quite long, of the order of μs – ms . Figure 40.2 shows some of the observed transitions in the case of the Pr^{3+} ion. There are further constraints on possible transitions which arise from an analysis of the angular momenta of the initial and final states. For example, a transition between two states both of which have $J = 0$ is forbidden since there is no angular momentum change as required for a photon; similarly for dipole transitions we require $\Delta J = 0, \pm 1$.

The transition probability per second for spontaneous emission [40.6] is given by

$$P_{ij} = \frac{64\pi^4\nu^3}{3hc^3} |r_{ij}|^2, \quad (40.3)$$

where ν is the frequency of the transition, h is Planck's constant, c is the velocity of light, and $|r_{ij}|$ is the matrix element of the electric dipole operator \mathbf{er}_{ij} between the two states i and j , and e is the electronic charge. For absorption, this must be multiplied by N , the mean number of photons with energy $h\nu$, which thus incorporates the effect of the incident beam intensity. Experimentally, one measures an absorption coefficient k as a function of energy $k(E)$ [40.6] which is linked to P_{ij} through,

$$\int k(E)dE = N_i \left(\frac{\pi e^2 h}{nmc} \right) \left(\frac{n^2 + 2}{3} \right)^2 f_{ij}, \quad (40.4)$$

where n is the refractive index of the crystal environment, m is the electronic mass, N_i is the concentration of the luminescent centres, and f is the *oscillator strength* for the transition. For both absorption and emission, the dimensionless oscillator strength f_{ij} defined as [40.6]

$$f_{ij} = \frac{8\pi^2 m \nu}{3h e^2} |r_{ij}|^2 \quad (40.5)$$

is often quoted to compare the relative transition probabilities. For an electron harmonic oscillator, $\sum f_{ij} = 1$, and so oscillator strengths of the order of 0.1–1 are strongly allowed transitions.

40.1.2 Transition-Metal Ions

Transition-metal ions from the 3d series are characterised by a much stronger interaction with the crystalline environment than the 4f ions since there is no equivalent of screening by the 5s, 5p outer shells. In addition, the spin-orbit coupling is weaker, and so the order

of perturbation is reversed: the atomic L , S multiplets are split by the crystal field, with spin–orbit coupling being a smaller interaction. Intra-configurational transitions are again strictly forbidden, but become weakly allowed by inter-configurational mixing through odd-parity crystal fields, and by odd-parity vibrations. The result of this is that the strongest selection rule after parity is that transitions should have $\Delta S = 0$, since the electric dipole operator does not involve spin. The other major difference, again due to the strength of the crystal-field interaction, is that transitions which are purely electronic, the so-called zero-phonon lines, are rarely observed. Rather what are seen are broad bands which correspond to the simultaneous excitation of an electronic transition and vibrational transitions, which overlap to give the broad observed bands. In particular, transitions involving odd-parity vibrations have a high transition probability through the effect of configuration admixing. This will be considered in a following section. The most commonly observed luminescent ions are those from the d^3 configuration (Cr^{3+} , Mn^{4+}) and from the d^5 configuration (Mn^{2+}).

40.1.3 s^2 Ions

The $5s^2$ (e.g. Sn^{2+} and Sb^{3+}) ions and $6s^2$ (e.g. Tl^+ , Pb^{2+} , Bi^{3+}) ions are of considerable importance because transitions to and from the excited s^1p^1 states are Laporte-allowed. The interaction of the p state with the

crystalline environment can be very strong, and so broad spectra are often observed.

40.1.4 Semiconductors

Luminescence in semiconductors is dominated by near-band-gap luminescence arising from recombination of electrons and holes. This process is most efficient in direct-band-gap materials such as ZnS and GaP rather than indirect-gap materials such as Si and Ge because the transition probability requires conservation of wavevector, but the photon wavevector is ≈ 0 on the scale of the Brillouin zone. Hence creation or destruction of a phonon is required for band-to-band luminescence in indirect-gap materials, which is less probable. The near-edge emission may correspond to luminescence from a variety of shallow energy-level structures such as free or trapped excitons, or from donor–acceptor recombination. These are both examples of electronic systems with spatially extensive wavefunctions, in contrast to the atomically localised $3d$ and $4f$ wavefunctions considered earlier. However, it is also possible to observe *deep-level* luminescence from transition-metal ions and rare earths in semiconductors provided that the electron affinities and band-gap energies are such that the pertinent energy levels fall in the band gap. Since semiconductors are discussed elsewhere in this volume, we shall not consider them further here.

40.2 Interaction with the Lattice

For rare-earth ions, the interaction with the vibrations of the crystal lattice can be ignored for most purposes; the observed luminescence spectrum consists of sets of sharp electronic transitions. But for other luminescent ions which interact strongly with the vibrating ions of the surrounding crystal, the incorporation of the latter is critical to explaining the observed spectra. The simplest model of ion–lattice interactions is to consider only the N nearest neighbour ions and their atomic displacements X_n , Y_n , Z_n , ($n = 1, N$) in Cartesian coordinates. The vibrational Hamiltonian involves cross terms in these coordinates, but may be transformed to harmonic form if symmetry-adapted forms of these coordinates (*normal modes*) are used instead of the actual displacements. For example, the so-called breathing mode Q_b , for an octahedrally coordinated ion takes the form

$$Q_b = (Z_1 - Z_2 + X_3 - X_4 + Y_5 - Y_6)/6. \quad (40.6)$$

If all the other modes have zero amplitude, the ions move radially towards or away from the central luminescent ion. The key point in considering the influence of the crystal lattice is that the vibrational potential energy is just the variation of the electronic energy with ionic displacement, or equivalently with the normal modes (within the spirit of the Born–Oppenheimer approximation). Put another way, the crystal field depends on the ion positions so that the electron and lattice quantum-mechanical systems are linked through this electron–lattice coupling. We can therefore expect a difference in the harmonic vibrational potential from one electronic state to another, so that it will in general have the form $(1/2)m\omega_i^2(Q - Q_{0i})^2$; i.e. both the magnitude of the potential and the position of the minima Q_{0i} will depend on the electronic state i . The vibrational states in the harmonic approximation are just the usual simple harmonic oscillator states with energies $(n + 1/2)h\nu_i$,

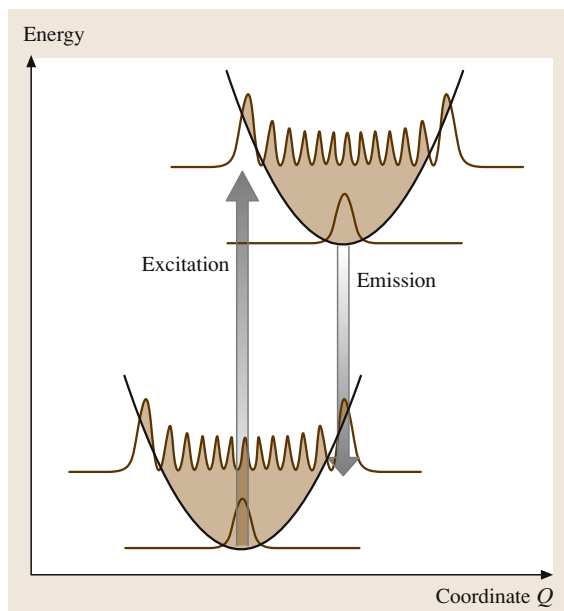


Fig. 40.3 Configuration coordinate diagram for excitation/emission cycle

with n an integer. Thus we arrive at a *configuration coordinate* diagram such as that shown in Fig. 40.3, with the potential energies for a ground state g and excited state e being offset parabolas of different curvature. The square of the vibrational wavefunction is shown for the ground state and that with $n = 11$. We note that for a classical oscillator there would be a peak in the probability function at the extreme lengths of travel, corresponding to the maxima for $n = 11$ at the outer limits of the wavefunction in Fig. 40.3.

We consider first luminescence from the excited electronic state/lowest vibrational state. The transition probability may be calculated using the Franck–Condon principle, that is we assume that the duration of the electronic transition is much shorter than a vibrational period, so that the Q remains constant during a transition. The transition is therefore taken to be *vertical* on the configuration coordinate diagram. The transition probability is just $\langle \phi_e | \phi_g \rangle^2$, the square of the vibrational wavefunction overlap, multiplied by the electronic transition probability considered earlier. From Fig. 40.3, this overlap will be a maximum for some vibrational state other than the ground state, unless the positions of the minima of the two potential energy curves coincide accidentally. (In Fig. 40.3, this maximum would be for the $n = 11$ ground vibrational state). Put another way, the maximum transition probability is not

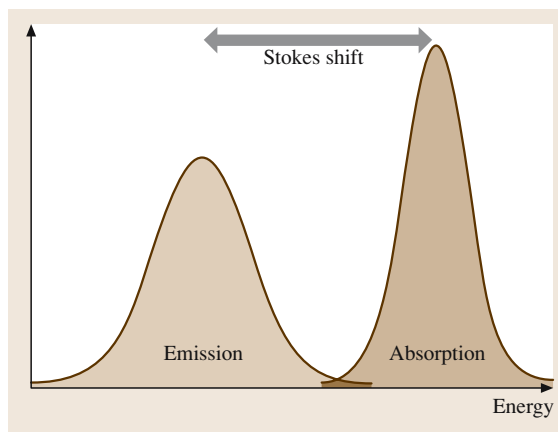


Fig. 40.4 Absorption and emission line shapes for strong electron–lattice coupling

for the zero-phonon transition (no change in vibrational state), but corresponds to the creation of a finite number of phonons. A range of transitions is allowed, and the result in a semiclassical analysis, allowing for finite line widths, is a Gaussian-shaped band. The analysis has to be extended to include finite temperatures and other modes, but the overall result is that the emission line shape is approximately a Gaussian centred at an energy lower than that of the difference between the minima of the two potential curves. The same argument can be applied to the excitation process, and again an approximately Gaussian line shape results but this time centred on an energy above that of the difference in potential-energy minima. Figure 40.4 shows the overall result; the difference between the maxima of the two curves, known as the Stokes shift, is an indicator of the degree of electron–lattice coupling.

It is clear from the diagram that the, in cases where there is strong electron–lattice coupling of this type, that: (a) there will be very little intensity in the zero-phonon line, and (b) there will be a large Stokes shift between the energies for maximum absorption and maximum emission. Thus the luminescence of transition-metal ions, colour centres, and closed-shell ground-state ions (s^2 , d^0), which have strong interactions with the lattice in the excited states, are typically broad bands with only occasionally weak, sharp zero-phonon lines being observed. For luminescence from within the 4f states of the rare earths, the reverse is true; we are in the weak-coupling regime, and zero-phonon lines are the predominant features of the spectrum.

40.3 Thermally Stimulated Luminescence

Thermally stimulated luminescence (TSL), or simply thermoluminescence (TL), refers to luminescence induced by thermally stimulated recombination of trapped electrons and holes in materials which have been subject to prior irradiation. The irradiation, which may be in the form of UV light, X-rays, gamma rays, or energetic electrons, creates free electrons and holes, most of which promptly recombine, but some of which are locally trapped at defect centres such as impurities and vacancies. If the trap binding energies are sufficiently large, thermal promotion of the electron or hole to the conduction band or valence band, respectively, is improbable at the irradiation temperature, and so these charge carriers remain trapped after irradiation. However, if the sample is then heated, thermally assisted recombination becomes increasingly probable, and the result is an initially increasing light output with increasing temperature until the traps are depleted, whereupon the light intensity drops. The curve of light intensity versus temperature $I(T)$ is known as a glow curve, and may be analysed to extract the trap depths and concentrations. A comprehensive review of the field has been given by *McKeever* [40.7], and a shorter discussion is given by *Vij* [40.1].

The process is shown schematically for a simple system comprising a single electron trapping level T and a single recombination centre R in Fig. 40.5.

Irradiation results in a trapped electron at trap T and a trapped hole at R . The trapped hole binding energy is larger than that of the trapped electron, so the latter is depopulated first, with a probability P which has the form,

$$P = s \exp(-E/kT), \quad (40.7)$$

where s and E are the attempt frequency and the activation energy respectively. We have shown the hole trap and recombination centre as being one and the same in Fig. 40.5, but it is also possible that the recombination energy is transferred to a separate luminescent centre.

In the measurement process, the sample is heated at a fast and linear rate, typically 1–10 K/s, whilst the light emission is monitored by a sensitive filter/photomultiplier combination, or by a monochromator system. The resulting glow curve may be fitted to a theoretical curve to extract the trap parameters. Generally the light output is quite weak, and TSL systems have sensitivity as one of their prime design factors, so cooled detectors and photon counting are commonplace. Above about 400 °C, thermal radiation from the sample/heater

is a problem and must be eliminated by filtering or by subtraction of a glow curve recorded using a thermally bleached sample.

The main uses of TSL are in determining trap depths and irradiation doses. In archaeological and geological applications, a comparison is made between accumulated natural dose and a dose from a radioisotope; by combining this with a measurement of the activity of the surroundings, a date since last thermal or optical erasure

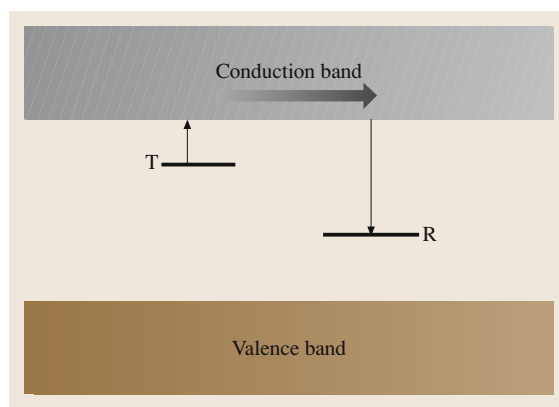


Fig. 40.5 Thermally stimulated luminescence process

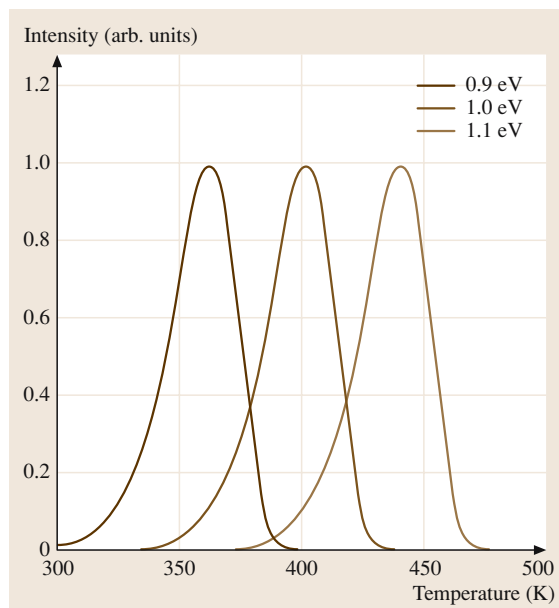


Fig. 40.6 Computed glow curves for first-order kinetics for fixed escape frequency and various trap depths

of the object can be determined. The nature of the traps is often poorly understood for these chemically complex samples. For medical dosimetry, room-temperature stable traps with high thermoluminescent output are required. At present, the material of choice is LiF doped with a few hundred ppm of Mg and Ti.

The mathematical form of the TSL glow curve depends on the physical model used for the TL process. In the simplest case, assuming first-order kinetics [40.1, 7], the light intensity I at temperature T is given by,

$$I(T) = n_0 s \exp(-E/k_B T) \times \exp \left[- (s/\beta) \int_{T_0}^T \exp(-E/k_B T) dT \right], \quad (40.8)$$

where n_0 is the number of occupied traps at time $t = t_0$ when the temperature is T_0 , β is the heating rate in K/s, and k_B is Boltzmann's constant. The result is a glow curve whose peak position varies approximately linearly with trap depth E , as shown in Fig. 40.6 (for fixed escaped frequency $s = 3.3 \times 10^{11} \text{ s}^{-1}$ and heating rate 1 K/s).

Of course, the peak position also depends on the escape frequency s , but is less sensitive to s than to E . For second-order kinetics [40.1, 7], the glow curve equation becomes

$$I(T) = \left(n_0^2 s / N \right) \exp(-E/k_B T) / \left[1 + (n_0 s / N \beta) \int_{T_0}^T \exp(-E/k_B T) dT \right]^2, \quad (40.9)$$

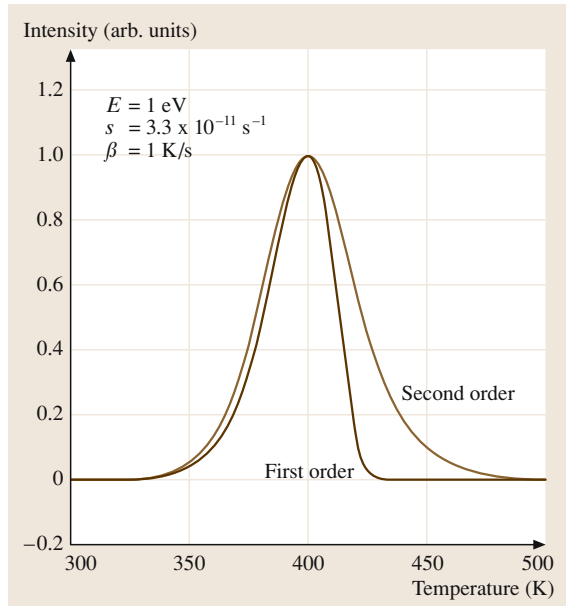


Fig. 40.7 Comparison of first- and second-order glow curves

where N is the number of available traps. The two forms both give glow curves of similar shape, but the first-order form shows an asymmetric form with a sharp fall off on the high-temperature side of the peak, whilst the second-order form is more symmetric as shown in Fig. 40.7. The parameters are extracted by least-squares fitting of these expressions to the experimental glow curves, which is cumbersome due to the integral, but *Kitis* et al. [40.8] have given analytical approximations to (40.8) and (40.9). In practice, many glow curves do not follow first- or second-order kinetics precisely.

40.4 Optically (Photo-)Stimulated Luminescence

Thermally stimulated luminescence is sometimes also accompanied by optically stimulated luminescence (OSL), in which one of the trapped carriers is excited by optical stimulation to a level from which it can recombine with the conjugate carrier by tunnelling, or completely to one of the bands so that recombination is achieved through what is essentially a photoconductivity effect. For OSL to be significant,

there must be an appreciable optical transition probability, so not all TSL centres are OSL active. The stimulation energy measured in OSL generally differs from that determined from TSL because of the Franck–Condon principle. The OSL effect finds practical application in dosimetry, e.g. [40.9], and in X-ray storage phosphors used for medical imaging, as described later.

40.5 Experimental Techniques – Photoluminescence

A typical traditional photoluminescence measuring system involves a broad-spectrum source, either a combined tungsten filament for the visible spectrum and deuterium lamp for the UV, or a xenon flash lamp. The lamp emission is passed through a grating monochromator and so selectively excites the luminescence. Band-pass or band-edge filters are generally required to eliminate unwanted second- and higher-order diffraction maxima from the grating. The luminescence is efficiently gathered by low- f /number optics and fed to a second grating monochromator, also equipped with filters, to monitor and analyse the luminescence. The final detector may be a photomultiplier, or preferably a charge-coupled device or photodiode array for improved data collection efficiency at multiple wavelengths.

This arrangement relies on good monochromatization/filtering to remove what is sometimes a relatively strong component of scattered light from the beam analysed by the emission monochromator. An alternative method of removing scattered light is to use time discrimination, by replacing the source by a xenon flash lamp (for example, as in the common Perkin Elmer LS55B luminescence spectrometer). The flash has a duration of about 10 μ s, so any scattered light has decayed away to an insignificant level when the emitted beam is sampled some time (0.1–10 ms) after the flash. A timed electronic gate is used to sample the emission immediately after a flash and just before the next flash; the difference between these two sampled signals gives the short-term luminescence whilst the second sample alone gives the long-term luminescence with *long* and *short* being relative to the flash repetition period. Of course, luminescence with lifetimes shorter than the pulse width ($\approx 10 \mu$ s) cannot be readily measured with this system. The luminescence intensity is normalised with respect to the excitation intensity by steering a sample of the excitation beam to a rhodamine dye cell which has a quantum efficiency of essentially unity for wavelengths below about 630 nm. The fluorescence from the dye is measured with a second photomultiplier.

To minimise the effect of scattered light, a conventional laser with its intrinsically narrow linewidth and high intensity is a very convenient replacement for a broad-spectrum lamp, but suffers from the disadvantage of a fixed wavelength. Typical lasers of interest are nitrogen (pulsed), argon (UV lines, or frequency-doubled visible lines), krypton, and the new

generations of GaN/GaInN blue/violet/UV laser diodes. For rare-earth spectroscopy, or other systems which are characterised by narrow absorption lines, it is very useful to have a scanning dye laser as the excitation source. This *selective excitation* facility enables *tagging* of particular luminescent levels with excited states belonging to the same centre, so that a picture of the energy-level structure of each luminescent centre can be built up in cases where several such centres contribute to the overall luminescence.

For decay kinetics on faster time scales, fluorimeters such as those developed initially by the Spex company (now Horiba) use a fast modulator and phase-sensitive detection to measure the phase shift between fluorescence and excitation; it is claimed that fluorescence decays can be measured with a resolution of 25 ps this way. An alternative is the time-correlated single-photon-counting technique which can measure decay constants in the ps–ns range. In this method, the excitation comes from a fast laser pulse, and the light level reaching the photomultiplier or micro-channel plate detector is reduced to such a low level that less than one photon per excitation pulse is detected. The time delay between the photon detection and the time of the pulse is measured, and a histogram produced of numbers of detected photons versus arrival time taken over a large number of excitation pulses. For efficient data collection, a high repetition rate and fast-pulse laser are required, often a Ti–sapphire laser. The wavelength for Ti–sapphire is too large for stimulating many materials directly with single-photon excitation, but stimulation is nonetheless possible by a two-photon excitation process, or by the use of a nonlinear crystal acting as a frequency doubler to produce laser output at one half the wavelength of the basic laser.

There are a number of specialist techniques, such as hole-burning, fluorescence line-narrowing, and photon echo methods associated with the use of lasers with either very narrow line widths or short pulse duration which have developed in a parallel way to techniques first introduced in nuclear magnetic resonance, and which are mainly used to investigate the dynamics and quantum mechanics of the luminescent species rather than the material in which they are contained. *Meijerink* gives a review of experimental luminescence techniques [40.1] which includes a short discussion of these specialist techniques.

40.6 Applications

The largest market for luminescent materials has traditionally been in the areas of lighting, through fluorescent tubes, and in cathode-ray-tube screen phosphors for image display. Both of these areas can now be regarded as mature in terms of materials development. However, new discoveries in the past decade, and the advent of new technologies, have rekindled interest in phosphor materials. Some of the current areas of activity in applications are outlined below.

40.6.1 White Light-Emitting Diodes (LEDs)

The development of blue, violet and UV LEDs based on GaN, InGaN and other semiconductors and alloys has stimulated great interest in the possibility of producing a *white-light LED* for use in lighting applications. LEDs are now available with emissions which peak as low as 365 nm in the ultraviolet. The concept is to use the blue emission to stimulate luminescence from yellow, or red and green phosphors, so that when mixed with residual blue light from the LED the result is simulated white light. For UV LEDs the LED output is used to stimulate blue, red and green phosphors.

The first generation of white LEDs from companies such as Nichia relied on a YAG:Ce phosphor to convert

some of the emission from a 465-nm GaInN LED into an orange/yellow emission centred at 550 nm; the combination of blue and yellow simulates white light. More recent versions of this scheme include $\text{Sr}_3\text{SiO}_5:\text{Eu}^{2+}$, which emits at 570 nm, and is claimed [40.10] to be more efficient than YAG:Ce, and a $\text{CaSiAlON}:\text{Eu}$ ceramic phosphor which offers improved thermal stability [40.11]. Improved colour rendition is obtained by using more phosphors and better balance between the various emissions, for example by replacing the YAG:Ce with $\text{SrGa}_2\text{S}_4:\text{Eu}$ (green) and $\text{SrY}_2\text{S}_4:\text{Eu}$ (red). Most recent developments based on 375 nm UV LEDs have used multiple emissions to achieve even better colour balance; for example Kim et al. [40.12] use $\text{Sr}_3\text{MgSi}_2\text{O}_8:\text{Eu}^{2+}$, Mn^{2+} , which has blue (Eu^{2+}), yellow (Eu^{2+}) and red emissions (Mn^{2+}).

A second development in this area is the substitution of another semiconductor for the phosphor – the so-called photon recycling technique. In this method, a layer of AlGaInP is used to absorb some of the blue incident radiation and down-convert it to the complementary colour. The advantage is that the fabrication/integration process is simpler compared to combining phosphor and semiconductor technology.

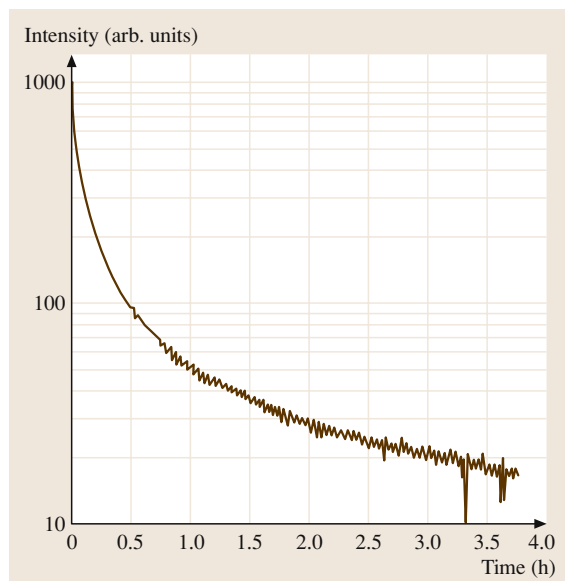


Fig. 40.8 Dark decay of persistent luminescence in a commercial lighting strip based on Nemoto $\text{SrAl}_2\text{O}_4:\text{Eu}/\text{Dy}$ material

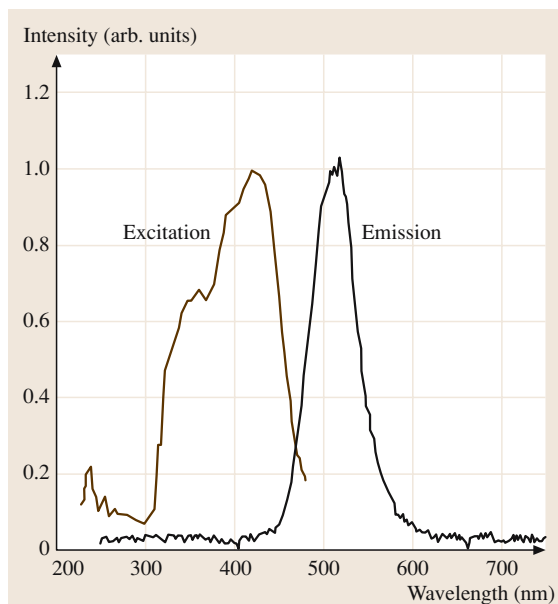


Fig. 40.9 Excitation and emission spectra of a commercial lighting strip based on Nemoto $\text{SrAl}_2\text{O}_4:\text{Eu}/\text{Dy}$ material

40.6.2 Long-Persistence Phosphors

It has been discovered in the past decade that some materials show a very long-lived but intense afterglow, arising from thermal emission of charge carriers from deep traps, followed by electron–hole recombination at or near a luminescent ion, i.e. room-temperature thermoluminescence. The persistence is for periods of several hours, much longer than the well-known ZnS:Cu persistent phosphor. This new class of materials have widespread applications in areas such as signage and passive emergency lighting for public buildings, aircraft cabins etc. The material is activated or *charged* by the blue/UV content of solar radiation or fluorescent indoor lighting during normal conditions; in subsequent dark conditions the energy is released as an afterglow. No power supply is required, the operation is entirely passive, and so a high degree of reliability is assured. Generally the phosphor powder ($\approx 10\ \mu$ grain size) is mixed with a resin binder and applied as a thick ($\approx 1\text{ mm}$) coating. The materials which have been used are rare-earth-doped strontium or calcium aluminate [CaAl_2O_4 :Eu/Nd (blue), SrAl_2O_4 :Eu/Dy (green), $\text{Sr}_4\text{Al}_{14}\text{O}_{25}$:Eu/Dy (blue–green), and $\text{Y}_2\text{O}_2\text{S}$:Eu/Mg/Ti (orange)]. Figures 40.8 and 40.9 show the decay and emission/excitation characteristics of a commercial phosphor strip of this type, based on Luminova® SrAl_2O_4 :Eu/Dy powder sourced from the Nemoto company. The decay is clearly non-exponential, but after the initial rapid decay the intensity may be fitted with an exponential with a decay constant of 3.1 h.

The traps which are responsible for the long-lived decays in these systems have not been clearly identified, and it is likely that there are several involved. Jia et al. [40.13, 14] have studied persistent violet/UV and blue ($\approx 400\text{ nm}$, 450 nm) luminescence from rare-earth ions in CaAl_2O_4 and BaAl_2O_4 and report that the thermally stimulated luminescence from these materials contains multiple glow peaks. They also describe how co-doping the calcium aluminate with Tb^{3+} results in a green persistent phosphor as a result of energy transfer from the cerium to the terbium ion; this mechanism could be the basis for convenient colour control, including white persistent phosphors. Aitasalo et al. [40.15] report that the nature of the traps is affected by the particular rare-earth ion or couple.

40.6.3 X-Ray Storage Phosphors

Optically stimulated or photo-stimulated luminescence is the basis for an X-ray imaging technology known

commercially as computed radiography (CR). CR was the first of a number of imaging techniques which are steadily replacing traditional photographic film methods. Several other techniques such as those based on amorphous selenium photoconductor/flat panels and a-Si arrays, and scintillator CCD/complementary metal–oxide–semiconductor (CMOS) arrays have also emerged in recent years, but CR is still the dominant of the new technologies on the basis of numbers of installed units. The many advantages of CR over photographic film, and the principles of the method are detailed in [40.16, 17], but the basic mechanism is that incident X-rays create electron–hole pairs in the material. Most of these promptly recombine, but some are trapped at defects and impurities, and remain trapped for periods of hours after the X-ray source is turned off. In an X-ray storage phosphor (XRSP), one of the trapped carriers is optically stimutable, and can be excited to the conduction band or valence band, or to a level from which it can recombine by tunnelling with the conjugate trapped carrier. The resulting recombination energy is transferred to a luminescent ion, and the intensity of the photo-stimulated luminescence (PSL) is in direct proportion to the incident X-ray intensity. The dominant material used in current XRSP systems is $\text{BaFBr}_{1-x}\text{I}_x\text{:Eu}^{2+}$, where the electron traps are F-centres, the hole traps are unidentified, and the luminescence is the $5d\text{--}4f$ transition of the Eu^{2+} ion. One disadvantage of systems based on this (powder) phosphor is that when the image is extracted with a raster-scanned He–Ne laser beam, light scattering from the powder grains means that material outside the focussed laser spot is also stimulated, limiting the spatial resolution to around $200\ \mu\text{m}$, which is inadequate for applications such as mammography. Several ways to overcome this are currently under development. RbBr:Eu is also an X-ray storage phosphor [40.18] and can be grown by vapour deposition in a columnar form.

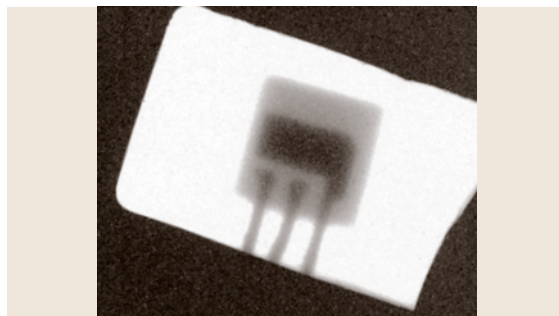


Fig. 40.10 PSL image of a BC549 transistor recorded on a glass-ceramic imaging plate

The columnar structure has a light-guiding property, restricting the scattering effect, and improving the resolution. A second development is that of glass-ceramic storage phosphors, where PSL active crystals are embedded in a glass [40.19–21]; the combination of particle size, separation and refractive-index mismatch means that these composite materials are semitransparent and the problem of scattering of read-out light is reduced. Figure 40.10 shows an image of a BC549 transistor recorded in a glass-ceramic X-ray storage phosphor.

40.6.4 Phosphors for Optical Displays

There are several new technologies being developed to replace cathode ray tubes for domestic televisions, including plasma display panels (PDPs). In these units, each pixel is a sealed cell containing a mixture of Xe and Ne in a dielectric-shielded electrode structure (for a review, see Boeuf [40.22]). An alternating current (AC) voltage applied between the electrodes results in a glow discharge being set up in the gas, and a Xe dimer vacuum UV (VUV) emission predominantly between 147 and 190 nm occurs. (In comparison, the mercury discharge in a conventional fluorescent tube emits primarily at 254 nm.) The UV discharge excites red, blue, or green phosphors coated on one of the cells; each colour is activated by an adjacent electrode. The requirements for efficient output from these phosphors differ from conventional tubes since the latter were chosen on the basis of their luminescence efficiency at a 254-nm pump wavelength, and their resistance to degradation by the UV light and chemical attack by Hg^+ ions. These requirements are evidently different for the PDP technology; in addition the phosphors must have a significant reflection coefficient in the visible to optimise the light output [40.23], and the surface quality is of greater significance due to the short penetration depth of the VUV. The phosphors which have been used so far include $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ (blue), $\text{Zn}_2\text{SiO}_4:\text{Mn}^{2+}$ (green), and $(\text{Y}, \text{Gd})\text{BO}_3:\text{Eu}^{3+}$ and $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ (red). The blue phosphor is prone to degradation.

The widespread introduction of Xe excimer excitation in PDPs can be expected to stimulate applications in other lighting technologies. In this regard, the possibility of so-called *quantum cutting* is of much interest. This recognises that the energy of a VUV photon is equivalent to two or more visible photons, so that quantum efficiencies in excess of 100% can in principle be achieved. The difficulty lies in finding a luminescent ion system whose energy-level system provides for both efficient pumping and two-photon luminescence in the visible. One

example which has been reported to have a quantum efficiency of up to $\approx 145\%$ is Pr^{3+} in YF_3 and other hosts [40.24, 25], where the excitation is through the allowed $4f^2 \rightarrow 4f^1 5d^1$ or host transitions. The system then decays to the $^1\text{S}_0$ excited state of the $4f^2$ configuration from which two-photon decay is possible through successive $^1\text{S}_0 \rightarrow ^1\text{I}_6$ and $^3\text{P}_0 \rightarrow ^3\text{H}_J, ^3\text{F}_J$ transitions, as shown in Fig. 40.2. (The intermediate step from $^1\text{I}_6$ to $^3\text{P}_0$ is provided by a nonradiative transition.) A difficulty is that the photon for the first transition is in the UV region of the spectrum, and so it is necessary to incorporate a second luminescent ion pumped by this transition to convert the UV to visible output, and the visible quantum efficiency is necessarily reduced.

40.6.5 Scintillators

Although semiconductor detectors of ionising radiation are making increasing inroads into the particle detection market, traditional scintillators are still widely used and are indispensable for some applications. The operating principle is that an incident gamma ray creates a large number of electron–hole pairs in the scintillating material directly or indirectly through the photoelectric effect, Compton scattering, or pair production, and that some of the energy of recombination appears as photon emission from luminescent ions. Charged particles such as protons produce electron–hole pairs through the Coulomb interaction with the band electrons. The *scintillation* or multiphoton bursts which signal the event is detected by a photomultiplier, and the height of the output pulse from the photomultiplier is proportional to the energy of the particle. A pulse-height analyser sorts the pulses according to energy, and so an energy spectrum can be obtained. Key figures of merit for a scintillator material are the numbers of photons per MeV of particle energy, the radiative lifetime of the luminescent ion (since possible pulse overlap limits the maximum count rate which can be measured), and the weighted density ρZ_{eff}^4 , which reflects the gamma sensitivity. (Here Z_{eff} is the effective atomic number.) A recent review of scintillators has been given by van Eijk [40.26]. The most widely used scintillator for many years has been $\text{NaI}:\text{Tl}$, but many different scintillators are being investigated, driven by the need for improved performance and lower cost for medical applications such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT), and for large-scale elementary-particle facilities such as those at the Centre Européen pour la Recherche Nucléaire (CERN). In the latter regard, the Crystal Clear collaborative project and other programs

have resulted in several new materials such as LaBr_3 , LaCl_3 , Lu_2SiO_5 , Gd_2SiO_5 , and LuAlO_3 , all Ce-doped, and undoped $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ and PbWO_4 , with typical performance figures of 10 000–50 000 photons/MeV and lifetimes of 10–50 ns. For gamma spectroscopy the pulse-height resolution is critical and $\text{LaBr}_3\text{:Ce}$ has twice the resolution of NaI:Tl . The fastest scintillators

are based on core-valence luminescence where a hole created in a core level recombines with an electron in the valence band. For example, BaF_2 shows this *cross-luminescence* effect with a lifetime as short as 600 ps. The effect is only shown by materials with a core-valence (CV) band energy gap less than the usual band gap, otherwise CV luminescence is absorbed.

40.7 Representative Phosphors

To conclude, we present in Table 40.2 a list of several luminescent materials of practical significance. The table is intended to be representative rather than comprehensive.

It is noticeable from the table that just a few ions are responsible for a large number of applications, and primarily as oxides.

Table 40.2 Some luminescent materials of practical significance

Host	Dopants	Colour	Excitation	Application
$\text{Bi}_4\text{Ge}_3\text{O}_{12}$	–	Blue	Ionising radiation	Scintillator
ZnS	Ag^+	Blue	Electrons	Colour TV screens
$\text{Zn}_{0.68}\text{Cd}_{0.32}\text{S}$	Ag^+	Green	Electrons	Colour TV screens
$\text{Y}_3\text{Al}_5\text{O}_{12}$	Ce^{3+}	Yellow	Blue, violet	White LED
Gd_2SiO_5	Ce^{3+}	UV	Ionising radiation	Scintillator
ZnS	Cu^+	Green	Electrons	Colour TV screens
BaFBr	Eu^{2+}	UV/blue	X-rays	X-ray imaging
$\text{BaMgAl}_{10}\text{O}_{17}$	Eu^{2+}	Blue	UV	fluorescent lamps, plasma displays
Sr_3SiO_5	Eu^{2+}	Blue	UV	White LED
SrGa_2S_4	Eu^{2+}	Green	UV	White LED
SrAl_2O_4	Eu^{2+} , Dy^{3+}	Green	UV, violet	Persistent phosphor
CaAl_2O_4	Eu^{2+} , Nd^{3+}	Blue	UV, violet	Persistent phosphor
Y_2O_3	Eu^{3+}	Red	Electrons, UV	Colour TV screens, fluorescent lamps
Sr_2SiO_4	Eu^{3+}	Yellow	UV	White LED
$(\text{Y}, \text{Gd})\text{BO}_3$	Eu^{3+}	Red	UV	Plasma displays
Y_2O_3	Eu^{3+}	Red	UV	Plasma displays
SrY_2S_4	Eu^{3+}	Red	UV	White LED
LiF	Mg^{2+} and Ti^{4+}	UV/blue	Ionizing radiation	TL dosimetry
ZnS	Mn^{2+}	Yellow	Electric field	Panel displays
Zn_2SiO_4	Mn^{2+}	Green	UV	Plasma displays
$\text{CeMgAl}_{11}\text{O}_{19}$	Tb^{3+}	Green	UV	Fluorescent lamps

References

- 40.1 D. J. Vij: *Luminescence of Solids* (Plenum, New York 1998)
- 40.2 G. Blasse, B. C. Grabmeier: *Luminescent Materials* (Springer, Berlin, Heidelberg 1994)
- 40.3 G. H. Dieke: *Spectra and Energy Levels of Rare Earth Ions in Crystals* (Interscience, New York 1968)
- 40.4 W. T. Carnall, G. L. Goodman, K. Rajnak, R. S. Rana: *J. Chem. Phys.* **90**, 343 (1989)

- 40.5 E. Merzbacher: *Quantum Mechanics* (Wiley, New York 1970)
- 40.6 D. Curie: *Luminescence in Crystals* (Methuen, London 1962)
- 40.7 S. W. S. McKeever: *Thermoluminescence of Solids* (Cambridge Univ. Press, Cambridge 1985)
- 40.8 G. Kitis, J. M. Gomez-Ros, J. W. N. Tuyn: *J. Phys. D.* **31**, 2636–2641 (1998)
- 40.9 L. Botter-Jensen, S. W. S. McKeever, A. G. Wintle: *Optically Stimulated Luminescence Dosimetry* (Elsevier, Amsterdam 2003)
- 40.10 J. K. Park, C. H. Kim, H. D. Park, S. Y. Choi: *Appl. Phys. Lett.* **84**, 1647–1649 (2004)
- 40.11 S. Ken, O. Koji, K. Naoki, O. Masakazu, T. Daiichiro, H. Naoto, Y. Yominobu, X. Rong-Jun, S. Takayuki: *Opt. Lett.* **29**, 2001–2003 (2004)
- 40.12 P. L. Kim, P. E. Jeon, Y. H. Park, J. C. Choi, L. P. Park: *Appl. Phys. Lett.* **85**, 3696–3698 (2004)
- 40.13 D. Jia, R. S. Meltzer, W. M. Yen: *Appl. Phys. Lett.* **80**, 1535–1537 (2002)
- 40.14 D. Jia, X. J. Wang, E. van der Kolk, W. M. Yen: *Opt. Commun.* **204**, 247–251 (2002)
- 40.15 T. Aitasalo, A. Durygin, J. Holsa, J. Niittykoski, A. Suchocki: *J. Alloys Comp.* **380**, 4–8 (2004)
- 40.16 S. Schweizer: *Phys. Status Solidi* **187**, 335–393 (2001)
- 40.17 J. A. Rowlands: *Phys. Med. Biol.* **47**, R123–R166 (2002)
- 40.18 P. Hackenschmied, G. Schierner, A. Batentschuk, A. Winnacker: *J. Appl. Phys.* **93**, 5109–5113 (2003)
- 40.19 S. Schweizer, L. Hobbs, M. Secu, J.-M. Spaeth, A. Edgar, G. V. M. Williams: *Appl. Phys. Lett.* **83**, 449–451 (2003)
- 40.20 M. Secu, S. Schweizer, A. Edgar, G. V. M. Williams, U. Rieser: *J. Phys. C: Condens. Matter* **15**, 1–12 (2003)
- 40.21 A. Edgar, G. V. M. Williams, S. Schweizer, M. Secu, J.-M. Spaeth: *Curr. Appl. Phys.* **4**, 193–196 (2004)
- 40.22 J. P. Boeuf: *J. Phys. D* **36**, R53–R79 (2003)
- 40.23 H. Bechtel, T. Juestel, H. Glaeser, D. U. Wiechert: *J. Soc. Inform. Display* **10**, 63–67 (2002)
- 40.24 S. Kuck, I. Sokolska, M. Henke, M. Doring, T. Schefler: *J. Lumin.* **102–103**, 176–181 (2003)
- 40.25 A. B. Vink, P. Dorenbos, C. W. E. Van Eijk: *J. Solid State Chem.* **171**, 308–312 (2003)
- 40.26 C. W. E. Van Eijk: *Nuclear Instruments and Methods in Physics Research A* **460**, 1–14 (2001)

41. Nano-Engineered Tunable Photonic Crystals in the Near-IR and Visible Electromagnetic Spectrum

Photonic crystals offer a well-recognized ability to control the propagation of modes of light in an analogous fashion to the way in which nanostructures have been harnessed to control electron-based phenomena. This has led to proposals and indeed demonstrations of a wide variety of photonic-crystal-based photonic devices with applications in areas including communications, computing and sensing, for example. In such applications, photonic crystals can offer both a unique performance advantage, as well as the potential for substantial miniaturization of photonic systems. However, as this review outlines, two-dimensional (2-D) and three-dimensional (3-D) structures for the spectral region covering frequencies from the ultraviolet to the near-infrared ($\approx 2\mu\text{m}$) are challenging to fabricate with appropriate precision, and in a cost-effective and also flexible way, using traditional methods. Naturally, a key concern is how amenable a given approach is to the intentional incorporation of selected defects into a particular structure. Beyond passive structures, attention turns to so-called active photonic crystals, in which the response of the photonic crystal to light can be externally changed or

- 41.1 **PC Overview** 998
 - 41.1.1 Introduction to PCs..... 998
 - 41.1.2 Nano-Engineering of PC Architectures..... 999
 - 41.1.3 Materials Selection for PCs 1000
- 41.2 **Traditional Fabrication Methodologies for Static PCs** 1001
 - 41.2.1 2-D PC Structures..... 1001
 - 41.2.2 3-D PC Structures..... 1007
- 41.3 **Tunable PCs**..... 1011
 - 41.3.1 Tuning the PC Response by Changing the Refractive Index of the Constituent Materials 1011
 - 41.3.2 Tuning PC Response by Altering the Physical Structure of the PC 1012
- 41.4 **Summary and Conclusions** 1014
- References** 1015

tuned. This capability has widespread potential in planar lightwave circuits for telecommunications, where it offers mechanisms for selective switching, for example. This review discusses alternative proposals for tuning of such photonic crystals.

Photonic crystals (PCs) are periodic, dielectric, composite structures in which the interfaces between the dielectric media behave as light-scattering centers. PCs consist of at least two component materials having different refractive indices, and which scatter light due to their refractive-index contrast. The one, two, or three-dimensional (1-D, 2-D, or 3-D) periodic arrangement of the scattering interfaces may, under certain conditions, prevent light with wavelengths comparable to the periodicity dimension of the PC from propagating through the structure. The band of forbidden wavelengths is commonly referred to as a *photonic band gap* (PBG). Thus, PCs are also commonly referred to as photonic-band-gap (PBG) structures.

PCs have great potential for providing new types of photonic devices. The continuing demand for photonic devices in the areas of communications, computing, and signal processing, using photons as information carriers, has made research into PCs an emerging field with considerable resources allocated to their technological development. PCs have been proposed to offer a means for controlling light propagation in submicron-scale volumes – the photon-based equivalent of a semiconductor chip – consisting of optical devices integrated together onto a single compact circuit. Proposed applications of PCs for the telecommunication sector include optical cavities, high-*Q* filters, mirrors, channel add/drop filters, superprisms and com-

pact waveguides for use in so-called planar lightwave circuits (PLCs).

Practical applications of PCs generally require man-made structures, as photonic devices are designed primarily for light frequencies ranging from the ultraviolet to the near-IR regime (i. e., ≈ 100 nm to ≈ 2 μ m, respectively) and PCs having these corresponding periodicities are not readily available in nature. 1-D PCs in this wavelength range may be easily fabricated using standard thin-film deposition processes. However, 2-D and 3-D PC structures are significantly more difficult to fabricate and remain among the more challenging nanometer-scale architectures to realize with cost-effective and flexible patterning using traditional fabrication methodologies. Recently, there has been considerable interest in PC-based devices that has driven advanced fabrication technologies to the point where techniques are now available to fabricate such complex structures reliably on the laboratory scale. In addition to traditional semiconductor nanostructure patterning methods based on advanced patterning/etching techniques developed by the semiconductor industry, novel synthesis methods have been identified for 2-D and 3-D periodic nanostructured PC arrays. There are several excellent reports reviewing these fabrication techniques in the literature [41.1–4] and this growing field has already

been the subject of numerous recent reviews, special issues, and books in the area of theoretical calculations (both band-structure and application simulations), 2-D PC structures, 3-D PC structures, and opal-based structures [41.5].

Recently, there has been great interest in exploring the use of PCs for active applications in the field of telecommunications, such as in the area of PLCs (e.g., for optical switching). In such applications, the PC properties should be adjustable to create *tunable* photonic band gaps. This development increases the functionality of all present applications of PCs by allowing the devices in such applications to be adjustable, or tunable. We review recent developments in the engineering of tunable nanometer-scale architectures in 2-D and 3-D. This review aims to organize this ever-changing volume of information such that interested theorists can design structures that may be easily fabricated with certain materials, and such that technologists can try to meet existing fabrication gaps and issues with current systems.

The chapter begins with a brief introduction to PCs, followed by general criteria used to determine appropriate methodologies for the nano-engineering of tunable PCs. Finally, overviews of the most common fabrication methods for tunable 2-D and 3-D PCs will be given.

41.1 PC Overview

41.1.1 Introduction to PCs

The simplest PC structure is a multilayer film, periodic in 1-D, consisting of alternating layers of material with different refractive indices (Fig. 41.1). Theoretically, this 1-D PC can act as a perfect mirror for light with wavelengths within its photonic band gap, and for light incident normal to the multilayer surface. 1-D PCs are found in nature, as seen for example in the iridescent colors of abalone shells, butterfly wings and some crystalline minerals [41.6], and in manmade 1-D PCs (i. e., also known as Bragg gratings). The latter are widely used in a variety of optical devices, including dielectric mirrors, optical filters and in optical fiber technology.

The center frequency and size (i. e., frequency band) or so-called stop band of the PBG depends on the refractive index contrast (i. e., n_1/n_2 , where n_1 and n_2 represent the refractive index of the first and second material, respectively) of the component materials in the system. Figure 41.1 is an example of a 1-D PC, with a pe-

riodic arrangement of low-loss dielectric materials. This multilayer film is periodic in the z -direction and extends to infinity in the x - and y -directions. In 1-D, a photonic band gap occurs between every set of bands, at either the edge or at the center of the Brillouin zone – photonic band gaps will appear whenever n_1/n_2 is not equal to unity [41.7]. For such multilayer structures, corresponding photonic band gap diagrams show that, the smaller the contrast, the smaller the band gaps [41.7]. In 1-D PCs, if light is not incident normal to the film surface, no photonic band gaps will exist. It is also important to note that, at long wavelengths (i. e., at wavelengths much larger than the periodicity of the PC), the electromagnetic wave does not probe the fine structure of the crystal lattice and effectively sees the structure as a homogeneous dielectric medium.

The phenomena of light waves traveling in 1-D periodic media was generalized for light propagating in any direction in a crystal, periodic in all three dimensions, in 1987 when two independent researchers suggested that

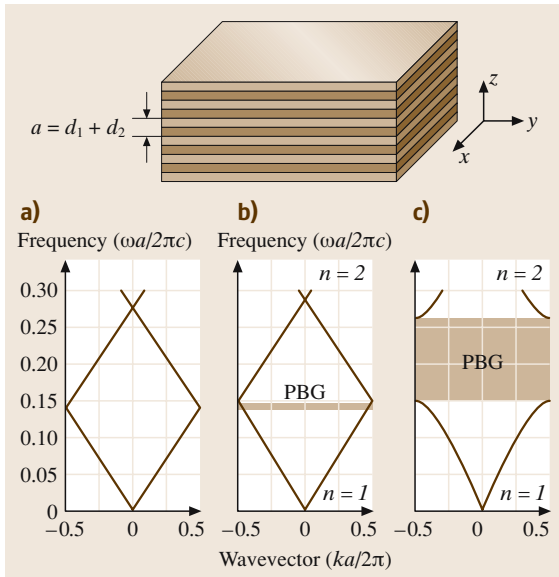


Fig. 41.1a–c Example of a 1-D photonic crystal. For a structure with $d = 0.5a$, the corresponding band gap diagrams are shown for: **(a)** GaAs bulk ($\epsilon = 13$); **(b)** GaAs/GaAlAs multilayer ($\epsilon_1 = 13$, $\epsilon_2 = 12$); and **(c)** GaAs/air multilayer ($\epsilon_1 = 13$, $\epsilon_2 = 1$) (from [41.7])

light propagation in 3-D could be controlled using 3-D PCs [41.8,9]. By extending the periodicity of the 1-D PC to 2-D and 3-D, light within a defined frequency range may be reflected from any angle in a plane (in 2-D PBG structures) or at any angle (in 3-D PBG structures).

Since the periodicity of PCs prevents light of specific wavelengths (i. e., those within the photonic band gap) from propagating through them in a given direction, the intentional introduction of *defects* in these structures allows PCs to control and confine light. Propagation of light with wavelengths that were previously forbidden can now occur through such defect states located within the photonic band gap. Defects in such PCs are defined as regions having a different geometry (i. e., spacing and/or symmetry) and/or refractive-index contrast from that of the periodic structure. For example, in a 2-D PC consisting of a periodic array of dielectric columns separated by air spaces, a possible defect would include the removal of a series of columns in a line. Specific wavelengths of light forbidden from propagating through defect-free regions would then be able to propagate through the line defect, but not elsewhere. Indeed, by appropriately eliminating further columns, light may be directed to form optical devices, including, for example, a low-loss 90° bend in a 2-D

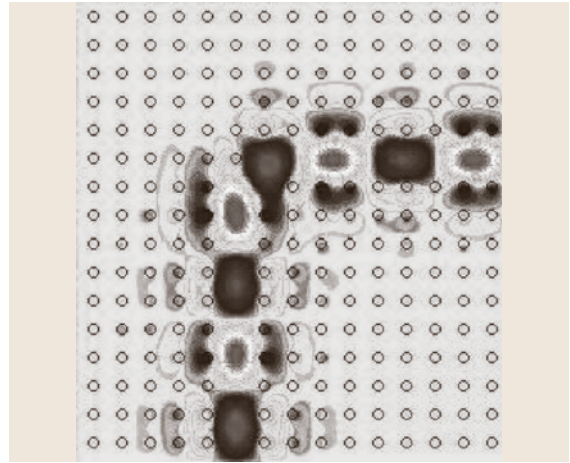


Fig. 41.2 Theoretical simulation of a low-loss 90° bend in a 2-D waveguide [41.7]

waveguide, as shown by theoretical simulation [41.7] (Fig. 41.2).

Clearly, photons controlled and confined in small structures, with size on the order of the wavelength of light, using the extremely tight bend radii offered by PCs would facilitate miniaturization and the fabrication of PLCs [41.7]. In addition, since the periodicity of the PC gives rise to the existence of band gaps which change the dispersion characteristics of light at given frequencies, defect-free PCs give rise to other interesting phenomena, including highly dispersive elements, through the so-called superprism effect [41.10]. Possible designs of PC-based optical devices using such properties have been extensively explored [41.1, 11], generating much excitement in the field of optical telecommunications [41.6].

41.1.2 Nano-Engineering of PC Architectures

Most of the promising applications of 2-D and 3-D PCs depend on the center frequency and frequency range of the photonic band gaps. A so-called *complete, full, or true* PBG is defined as one that extends throughout the entire Brillouin zone in the photonic band structure – that is, for all directions of light propagation for photons of an appropriate frequency. An incomplete band gap is commonly termed a *pseudo-gap* or a *stop band*, because it only occurs for reflection/transmission along a particular propagation direction. A complete gap occurs when stop-band frequencies overlap in all directions in 3-D. The center frequencies and stop-band

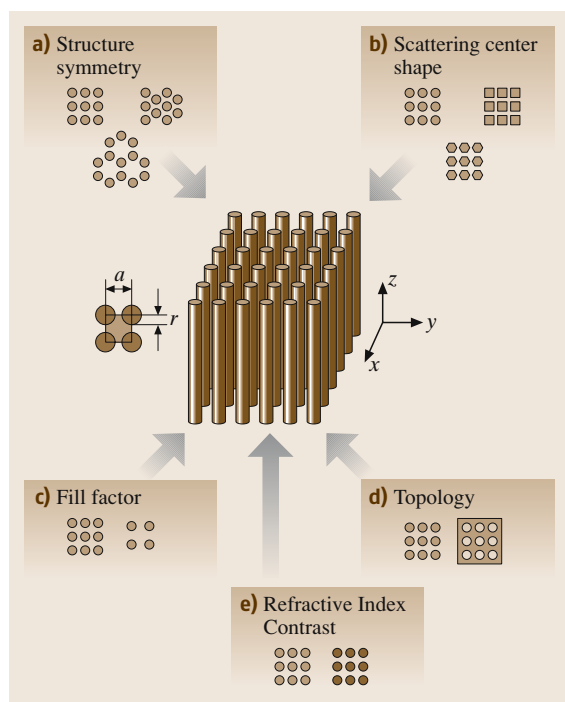


Fig. 41.3 PBG parameters that affect the center frequency and associated stop band (shown in *top view*, in *boxes*) relative to the 2-D periodic arrangement of one material (i. e., the rods, in *grey*) embedded in a second material (i. e., air), periodic in the x - y plane (shown in the *central figure*)

locations of the PBGs critically depend on the unit-cell structure [41.7–9, 12, 13]. In particular, the PC properties depend on the symmetry of the structure (i. e., the unit-cell arrangements), the scattering element shape within the unit cell, the fill factor (i. e., the relative volume occupied by each material), the topology, and the refractive-index contrast (Fig. 41.3).

It has been shown that a triangular lattice with circular-cross-section scattering elements in 2-D, or a face-centered-cubic/diamond lattice with spherical scattering elements in 3-D, tend to produce larger PBGs [41.7]. Also, as discussed above, the dielectric contrast is an important determining factor for 2-D and 3-D PC structures. The lower the dimension of the structure, the more readily a PBG manifested, since overlap of the PBGs in different directions is more likely (something which is a certainty in 1-D structures). In 3-D structures, calculations have determined that the minimum dielectric contrast (n_1/n_2) required to obtain a full PBG is about 2.0 [41.2, 12, 14]. For full PBGs, the ideal structure typically consists of a dielectric–air combina-

tion, to obtain both the greatest dielectric contrast as well as to reduce losses associated with light propagation in optical materials other than air [41.7].

The relationship between the fill factor, structure symmetry, and scattering element shape on the size and location of the PBG is complex [41.7, 15–18] and will not be discussed in detail here. However, it is clear that the ability to adjust, or tune, one or more of these parameters and thus tune the PC properties, is a very exciting development for future PC device applications.

41.1.3 Materials Selection for PCs

The physical architecture of a given PC is just one of the design considerations – another important one is the optical properties of the PC materials. In particular, the refractive index of a given material and its electronic band gap determine the performance and appropriate range of frequencies of PC devices fabricated from such a material.

For large PBGs, component materials need to satisfy two criteria; firstly, having a high refractive-index contrast, and secondly, having a high transparency in the frequency range of interest. This may be a challenge to satisfy at optical wavelengths. Suitable classes of PC materials include conventional semiconductors and ceramics, since at wavelengths longer than their absorption edge (or electronic band gap), they can have both high refractive indices and low absorption coefficients. In addition, these materials have other very useful electronic and optical properties that may complement the functions served by the presence of the PBG.

Another, often overlooked, consideration for PC material selection is the ability to translate the desired bulk single-crystal properties of well-known materials into nanometer-scale PC properties. Such structures typically have extremely large surface areas, and the microstructure of the constituent PC elements must be controlled during fabrication. Consequently, the final properties of the PC elements often vary from those of the bulk properties. This is extremely relevant when considering the functionality of a final PC device, such as those relying on electronic properties (e.g., lasing).

Finally, it is interesting to note that there is a strong correlation between the fabrication methods selected and the component materials. For example, top-down dry etching of semiconductors (e.g., Si, GaAs, InP, etc.) into nanostructured 2-D arrays is well characterized and relatively commonplace, whereas the same cannot be said for ceramic materials. The opposite is generally true for chemical or bottom-up synthesis using sol–gel

technology, in which ceramic PCs are relatively easy to fabricate when sol-gel-based infiltration techniques are used (e.g., in 3-D inverse opal fabrication). It is also

clear that the fabrication technique primarily determines the PC structure that can be fabricated, which in turn, determines the PC properties.

41.2 Traditional Fabrication Methodologies for Static PCs

Unlike for electrons, where there are natural length scales, for photons there is no such length scale [41.19]. This lack of an absolute length scale ensures that the physics of PCs is scalable: resizing the system rescales the energy in such a way that the spectrum, in units of c/a (where c is the speed of light and a is the lattice parameter of the PC), is independent of the system size [41.19]. Since $\omega/(2\pi c/a) = a/\lambda$, it is customary to use such dimensionless units to measure the photon energy [41.19]. Besides spatial scaling, there is also dielectric scaling – that is, there is no fundamental scale for ϵ . Two systems whose dielectric functions scale by a factor, $\epsilon'(\mathbf{r}) = \epsilon(\mathbf{r})/s^2$ have spectra that are scaled by the same factor: $\omega' = s\omega$, which means that increasing the dielectric function by a factor of four decreases the photon energy by a factor of two [41.19].

The apparent scalability of PCs is misleading because it suggests that PCs are truly size-independent, which in practice is not the case. As the size of the PC elements decrease, the effect of structural inhomogeneities become increasingly important, leading to inherent size limitations for particular architectures and for particular materials systems. This is most evident in changes in effective refractive index (e.g., due to the formation of inter-granular phases, surface oxides, etc.) and functionality.

2-D systems are increasingly being used for micro-photonic applications and consist of periodic repetitions of objects in a 2-D arrangement [41.19]. 3-D systems present a modulation in the third direction, such as a stack of spheres [41.19]. To date, 2-D and 3-D PCs have been fabricated using various approaches, including those based on electron-beam lithography, self-assembly and templating. The stringency of the requirements for a fabrication methodology, are dictated by the application of a given PC, in conjunction with associated requirements for material and architecture. Currently, there remains a great need for fabrication methods for mass production of nanometer-scale PCs, offering flexibility in material composition and design, while overcoming the low throughput of serial lithographic patterning, and the low reproducibility and material

restrictions of self-assembly methods. Many excellent reports reviewing these fabrication techniques are available in the literature [41.1–4]. For readers unfamiliar with traditional PC fabrication techniques, the following sections review typical 2-D and 3-D PC fabrication methods, based on lithography, self-assembly and hybrid techniques.

41.2.1 2-D PC Structures

Introduction to 2-D PBG Structures

The advantage of the 2-D PC structure is that, due to its planar form, it is easier to fabricate using existing lithographic methods, it retains its original single-crystal properties, and it can be simply integrated into desired PLC designs. However, there are specific design limitations that must be considered for 2-D PC structures. The main problem that has been identified both theoretically and experimentally is diffraction losses in the third dimension (i.e., for light directed out of the plane of the 2-D PC) [41.1, 2].

Calculations often assume a perfect and infinite 2-D material. However, the aspect ratio in experimental structures fabricated using typical micromachining techniques (e.g., reactive-ion etching) tend to be very limited [41.2]. Thus, in order to confine light in the third dimension (i.e., the out-of-plane dimension), the structure must either be large compared to the beam size (i.e., many wavelengths deep), or must be sufficiently confined within a waveguide to experience the full interaction with the periodic lattice [41.1]. As a consequence, 2-D PC structures are now typically fabricated in three principal forms: (1) a 2-D array of air holes in a dielectric (or dielectric pillars of very high aspect ratio in air), (2) a slab waveguide consisting of a 2-D array of holes perforating a thin membrane (with the membrane thickness approximately equal to the hole diameter) and (3) a heterostructure waveguide (i.e., similar to those used for confining light and carriers in semiconductor laser heterostructures) with holes drilled through the heterostructure [41.20]. New fabrication methodologies need to be able to address the fabrication issues for each of these three types of low-loss 2-D PC structures in

order to implement these structures in practical device applications.

2-D Lithographic PC Fabrication Methods

The majority of semiconductor PCs are synthesized using the advanced lithographic and etching capabilities developed for the semiconductor microelectronic industry [41.1, 2, 4, 21]. One major benefit of being able to fabricate PCs from single crystals using top-down lithography is that the properties (e.g., nonlinear electro-optic properties) are typically superior to those for polycrystalline structures produced using bottom-up (e.g., template-based) processes [41.22]. Utilizing the former materials enables significant modulation of the optical properties suitable for developing tunable PC devices. Also, in top-down fabrication, the optical and dielectric properties of the bulk material may be characterized before PC fabrication, allowing them to be more effectively modeled after PC fabrication. Indeed, semiconductor-based PBG structures made using lithographic processes tend to be more reproducible than self-assembled PC structures. Most importantly, there is much greater flexibility in the design and implementation of lithographically fabricated structures, and hence to implementing functional devices. Finally, PC-based PLCs are ideal for optical system integration.

For high refractive-index contrast (i.e., larger photonic stop bands), the preferred PBG structure is a topologically patterned substrate (i.e., an air-solid PC), which may be fabricated using multistep submicron machining based on lithography. In this process, a radiation-sensitive resist is patterned to transfer structures to the bulk material. The patterned resist is developed (i.e., the properties of the exposed material are changed), followed by etching and/or metal thin-film deposition to define features in the bulk material. The majority of submicron machining processes involve exposure of the resist with either electromagnetic radiation (e.g., optical, UV or X-ray photons) or charged particles (electrons, low-energy heavy ions, high-energy light ions) [41.23]. Although resist exposure is essentially a surface micromachining technique and therefore planar, various wet and dry etching techniques have been successfully used to produce topologically varying microstructures in the axial direction [41.23].

Similar to all resist-based processes, the resist itself limits the ultimate minimum size and maximum density of nanostructured patterns [41.24, 25], and organic residues (from processes resist) are often sources of contamination [41.26]. Organic resists are chosen as a compromise between sufficient resilience for selected

postprocessing procedures, and having sufficiently high sensitivity and contrast (i.e., a low energy threshold for resist exposure and total resist exposure occurring only above a certain energy threshold with no resist exposure occurring below, respectively) to the incident exposure [41.27] – this tends to be a difficult compromise to satisfy in practice for traditional resist materials.

Lithographic methods provide a simple and straightforward approach for the intentional introduction of defects into these 2-D systems (due to the simple surface patterning), and this top-down method may be the only way to retain the bulk-like semiconducting properties that are required for certain functionalities (e.g., such as lasing).

Electromagnetic Radiation Patterning. Generally, using electromagnetic radiation to micromachine structures requires masks to expose a resist material in a spatially selective manner, with subsequent development of the exposed resist to produce microstructures. Masks typically incorporate multiple repeat patterns, enabling multiple microstructures to be fabricated in a single exposure, which significantly reduces costs [41.23]. Masks tend to be less useful for patterning with charged particles, since the high energy deposited into the mask during exposure can produce undesirable mask instabilities due to heat expansion, stress and damage [41.23]. Charged-particle micromachining is therefore normally limited to direct-write processes, where a focused charged-particle beam is scanned over a material in a specific pattern to produce microstructures. Such a direct-write process has the advantage of not requiring a mask, but suffers from relatively low throughput as a serial write process, and hence losses its attraction for multiple component production [41.23].

A main consideration of patterning approaches for PC fabrication is the minimum required feature size. Diffraction effects occurring for wavelengths of light below 100 nm impose a fundamental limitation on achievable structure sizes using optical lithography. Micromachined features smaller than about 250 nm are essentially beyond current readily available optical lithography systems [41.23], although recent advances (e.g., in UV lithography) will probably make this size regime achievable in the near future using a parallel patterning processes. Other techniques such as X-ray lithography that are capable of high-volume production, by allowing parallel exposure for larger-area nanometer-scale patterning, are hampered by challenges including making suitable high-resolution masks, as well as the significant capital cost required to achieve the high pho-

ton flux suitable for mass production of nanometer-scale components [41.23].

Charged-Particle Beams. Limitations on the minimum feature size using traditional photolithography have motivated a number of studies using serial patterning with charged-particle beams (e.g., electron-beam lithography (EBL) or, less commonly, focused ion-beam (FIB) lithography) to obtain the small feature sizes required in PC structures.

The advantages of EBL and FIB for defining sub-100-nm feature sizes in a mask-less fashion are offset by the fact that high-volume production is commercially impractical using such serial writing techniques [41.23, 28, 29]. Conventional ion implantation, though widely available and robust, has thus far been incapable of patterning densely organized nanostructures due to ion scattering [41.27, 30] and mask limitations [41.31–35].

Another consideration for choosing lithographic techniques for sub-100-nm-scale feature fabrication is that all the steps in this multistep process have their own limitations. Indeed, the finite size of the molecular compound comprising the resist sets the ultimate limit on pattern resolution [41.24, 36]. Since all traditional submicron micromachining techniques except FIB require resist exposure, this is a major factor for determining resolution – for example, the ultimate resolution of EBL is not determined by the electron optical system, which is ≈ 0.1 nm, but by the resolution of the resist and by the subsequent fabrication process [41.37]. Also, electrons in the beam scatter very easily from electrons in the resist material, and both scattered and secondary electrons cause a lateral spread in the patterned material due to diffusion effects (i.e., proximity effects) [41.23]. Thus, the very high spatial resolution achievable at the surface using EBL deteriorates as the beam penetrates into the resist, making EBL unsuitable for high-resolution topographically patterned structures using resist alone [41.23]. The advantage of mask-less processing for high-definition surface structures in resist and the inherently slower processing mean that EBL has been limited to mask-making and direct-writing on wafers for specialized applications such as fabrication of 2-D PCs for research purposes only [41.23].

Currently, the most commonly used method to fabricate 2-D PCs in the optical regime is EBL-based mask processing, combined with etching [41.1]. This approach has been used to pattern different semiconductor materials, including Si [41.4] and GaAs [41.1, 4, 7] (Fig. 41.4). Some detrimental effects arising from dry etching have been reported, including damage and ion

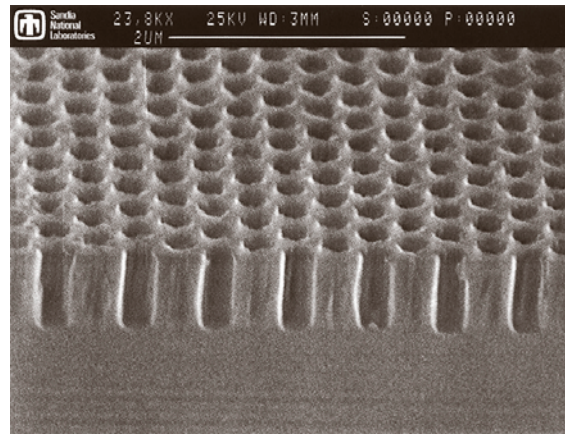


Fig. 41.4 Example of a 2-D PC structure in GaAs [41.7]. The air hole radius is 122.5 nm, the lattice parameter is 295 nm, and the holes are 600 nm deep

channeling that increase nonradiative recombination in active microstructures [41.1]. Other post-mask procedures used to fabricate 2-D PC structures include vertical selective oxidation and lift-off techniques [41.1]. In vertical selective oxidation, an *all solid* 2-D PC structure may be fabricated, which maintains the physical integrity of the crystal and allows deposition of electrical contacts afterwards. However, the oxidation process usually also occurs laterally, while typical desired configurations for 2-D PC crystals are structured in the vertical direction. Even with improved vertically selective oxidation techniques, the depth-to-diameter aspect ratios of pillars are around 1 : 1, which is significantly less than the 10 : 1 ratios achieved using dry etching techniques [41.1]. Also, volume shrinkage of up to 13% [41.1] can lead to strain, the formation of micro-cracks and other structural problems. 2-D PC membranes have been fabricated using epitaxial lift-off/substrate removal, but this method is limited to materials with sufficient wet-etch selectivity [41.1].

Despite its extremely low throughput, focused ion beams are presently used for direct, resistless, nanometer-scale surface modification, micromachining, and ion implantation. This is primarily because such patterning cannot be achieved with any other method. However, FIB patterning is inherently slow and expensive, and requires sophisticated optics for ion focusing. Also, the patterning resolution and proximity of nanometer-scale features is greatly limited by the significant FIB tail distribution [41.26, 36] and the highly dose-dependent sputtering yield of FIBs [41.38]. The use of FIB patterning has not been widely used to fabricate PCs.

Table 41.1 Summary of current submicron resist-based machining techniques [41.23]

Technique	Potential sub-100nm capability	Requires mask	Direct write	Facility availability	High-volume production capability	3-D capability
Optical lithography	No	Yes	No	Widespread industrial use	Yes	No ^a
X-ray lithography	Yes: depends on mask	Yes	No	Scarce and expensive	Yes	Yes
UV lithography	Yes: depends on mask	Yes	No	Scarce	Yes	Yes
Electron-beam writing	Yes: currently achieved	No	Yes	Medium availability	No	No
Low-energy ion beam writing	Yes: depends on focusing	No	Yes	Scarce	No	No ^b
High-energy ion beam writing	Yes: depends on focusing	No	Yes	Very scarce	No	Yes

^a Optical lithography is a surface technique. However, new supplementary dry etching techniques such as plasma etching are being used to produce submicron 3-D structures from 2-D surface structures in silicon

^b Low-energy ion-beam micro-machining (FIB) relies on sputtering atoms from the surface. 3-D structures can be produced by continuous erosion at a given location although, in practice, this is far too slow for practical 3-D microcomponent production.

Light ions collimated by stencil masks can be used to expose resists in a parallel fashion in both ion-projection lithography and masked ion-beam lithography [41.27]. The resist is exposed in a similar manner to X-ray and EBL. Light ions are generally preferred over heavier ions in ion lithography because they have the greatest range in resists (thick resists can be used to give low defect densities [41.31]), but protons also scatter more than heavier ions at the end of the implant range, limiting resolution and nanometer-scale feature proximity. However, the use of heavier ions to pattern, following the traditional approach used for semiconductor doping, can lead to mask erosion. This effect is negligible when light ions (e.g., hydrogen) are used, but can become pronounced for higher-dose implantation of heavier ions [41.27]. A summary of current lithography-based submicron micromachining techniques is given in Table 41.1.

2-D Self-Assembled PC Fabrication Methods

For 2-D PC structures, electrochemical methods (i.e., anodic etching and growth) and fiber-pulling have used to prepare extremely high-aspect-ratio ($\gg 50 : 1$) large-area nanometer-scale hole arrays (Fig. 41.5). These types of structures represent near-ideal infinite 2-D PC structures [41.20]. However, these methods are limited to specific materials (e.g., silicon [41.41] and alumina [41.39] for electrochemical methods, or glass [41.40] by fiber-pulling), and processes such as electrochemical etching are often limited in their capability to control fill factor, structure symmetry and scat-

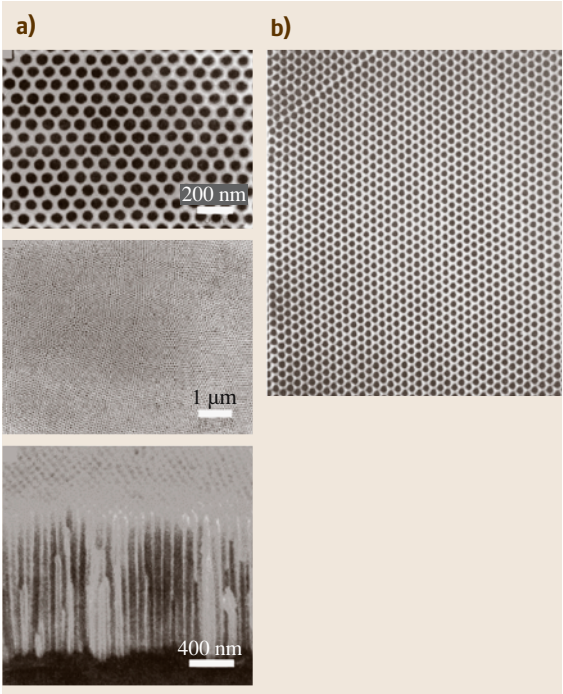


Fig. 41.5a,b Examples of 2-D self-organized PC structures of: (a) NCA [41.39], (b) NCG (nano channel glass) [41.40]

tering element shape. 2-D PC structures with long-range periodicity obtained by electrochemical etching often require surface pre-patterning by serial lithography to obtain lattice dimensions in the visible to near-IR range.

Table 41.2 Parameters of current high-aspect-ratio 2-D periodic structures [41.42–44, 46]

Nanotemplate type	Alumina	Silicon	Glass
Pore size (nm)	1–10 ²	400–10 ⁴	> 10
Pore density (cm ⁻²)	10 ⁹ –10 ¹²	10 ⁶ –10 ⁹	10 ¹⁰
Pore arrangement	Pseudo-ordered ^a	Disordered ^a	Ordered
Thickness (nm)	10 ³	10 ³	10 ³
Temperature tolerance	< 450 °C	< 900 °C	< 600 °C

^a May be ordered using pre-patterning processes [41.42, 45]

Some examples of easily available, high-aspect-ratio nanoporous arrays include electrochemically grown porous alumina (nanochannel alumina, or NCA) [41.39], electrochemically grown silicon [41.42] and fiber-pulled nanochannel array glass (Table 41.2) [41.43, 44].

Nanochannel Alumina. NCA may be fabricated economically with well-characterized properties. Pseudo-ordered regions can be easily fabricated using electrochemical etching, compared with porous silicon, which requires back-side illumination in addition to an applied bias. The pore arrangement is completely disordered in silicon if no pre-patterning with optical or EBL is used [41.46].

Aluminium is electrochemically oxidized to alumina (Al₂O₃) under positive polarization. For over a century, the growth of disordered pore arrangements have been observed and studied for selected electrolytes that weakly dissolve alumina. However, in 1995 it was first discovered that, after an extended period of anodization, self-ordered porous alumina films with pores arranged in a hexagonal pattern are obtained at the growth front [41.39]. Although knowledge of how to fabricate such NCA templates is now widespread, the formation mechanism of hexagonal NCA has still not been completely elucidated. Mechanical stress between neighboring pores due to the volume expansion of alumina with respect to the aluminium substrate has been proposed as a mechanism for the self-ordering [41.46].

Controlled single-domain porous arrays may be obtained by lithographically pre-patterning the aluminium substrate, and applying the optimum potential for the corresponding inter-pore distance [41.47]. Recently, in addition to the hexagonal lattice, square and honeycomb lattices with square or triangular pore shapes have also been obtained by appropriate pre-patterning [41.48].

Extremely high-aspect-ratio NCA templates can be fabricated with periodicities ranging from ≈ 80 to ≈ 600 nm. The domain sizes (i. e., areas with the same orientations) are controlled by the anodization condi-

tions and increase with time to micrometer-scale sizes. The thickness of the NCA template can be controlled by the electrochemical etching time, from ≈ 50 nm to > 10 μ m thick.

Current Alternative 2-D PC Fabrication Methods.

Templating of 2-D Structures. 2-D PCs have been fabricated by templating a secondary material into a PC backbone, that can either remain in the final structure as a composite material or be removed to form a dielectric/air PC structure. In many cases, a dielectric/air PC structure is preferred due to the higher dielectric contrast, but depending on the original PC backbone material, it may not be possible for the secondary materials to be removed after templating.

Over the past decade NCA templates have been commonly used to fabricate various nanostructured arrays by either dry etching or by growth through templates [41.45, 49, 50]. This method of using self-assembled templates for nanostructure patterning has been used to pattern physical topology, resulting in either nanostructured hole arrays (e.g., by dry etching through the templates), or growth of nanostructured wire arrays [e.g., by metalorganic chemical vapor deposition (MOCVD), MBE, and chemical-solution deposition]. The effectiveness of NCA as a templating tool results from the fact that the alumina may be easily selectively chemically etched compared with semiconductor materials, and forms a robust mask for dry etching (compared to, for example, patterned resist materials).

Combining Self-Organized 2-D Structures with Ion Implantation. Ion-beam techniques may be used to pattern, for example, sub-100-nm-resolution features on thin membranes, but they are unsuitable for deep sub-surface patterning, particularly at the end of the implantation range. One reason for this is that stencil-type masks need to be used, due to the relatively short range of energetic ions compared to, for example, X-rays, which do not require stencil-type masks [41.24, 27]. In the

past, high-aspect-ratio stencil masks for ion implantation were fabricated using techniques such as EBL, but such stencil masks were limited by the low pattern density achievable using such fabrication methods, requiring the use of self-complementary masks with multiple offset exposures to create printed patterns [41.34]. Also, the minimum resolution using nanoscale ion implantation through masks is so sensitive to the mask's wall angle [41.32, 33] that advanced photoresists and lithographic techniques need to be used to achieve such ideal masks.

Investigations have been reported on new ways to combine self-organized 2-D PC structures with conventional, broad-beam ion implantation to achieve nanoscale pattern transfer [41.51, 52]. Using self-organized templates as masks significantly simplifies the process, as no resolution-limiting resist or no time-intensive serial patterning is necessary, and these ideal, stencil-type masks allow direct nanostructure array transfer with significantly less scattering than solid masks. The use of self-organized masks allows this process to be economically feasible for large-area nanostructure array patterning. This approach transfers the desired pattern from the self-organized template into a desired material that cannot be patterned on the same scale using self-organization.

Despite concerns about mask edge defects [41.53–55], extended radiation damage accumulation extending below the mask edges [41.30, 35] and possible mask distortion from stress and temperature [41.31], it has been shown that nanoscale features could be retained using conventional ion implantation combined with high-aspect-ratio masks to make PC-type structures [41.51, 52] (Fig. 41.6). Energetic, charged particles were accelerated into a substrate using broad-area parallel implantation through NCA templates to pattern 2-D nanostructured arrays of ions in both semiconductor (i.e., InP) and ceramic (i.e., SrTiO₃) single crystals [41.51, 52]. The ion-implanted volume was selectively etched away after implantation to form 2-D nanostructured air-hole arrays in the substrate material. The array pattern was successfully fabricated over the entire implant area ($\approx 1 \text{ cm}^2$), with resulting minimum feature sizes of $\approx 40 \text{ nm}$. Such a direct and parallel method, combining ion implantation with a nanoscale mask, can be immediately beneficial, particularly in applications where using FIB is impractically slow (e.g., for substrate nanostructure array pre-patterning for selective or electrochemical etching). In addition, complex profiles (e.g., membrane formation) were shown to be possible using this technique. This method allows ions to

be implanted into nanoscale areas, which may result in additional substrate modification, including, but not limited to, micromachining of implanted areas to fabricate topologically patterned nanostructured arrays.

The ability to process structures of different form and shape using ion-implantation techniques offers opportunities for nano-machining, and enables material processing on a nanometer scale – this is applicable to materials which cannot be structured using standard lithography (e.g., samples with pronounced topography [41.38]). This nanoscale patterning method can also be used to directly pattern substrates to create electrochemical initiation sites for fabrication of ordered porous Si and porous GaAs structures for light emission [41.56] and for 3-D PC fabrication in Si [41.57]. In addition, such array pattern transfer is applicable to pre-patterning of self-assembled templates themselves,

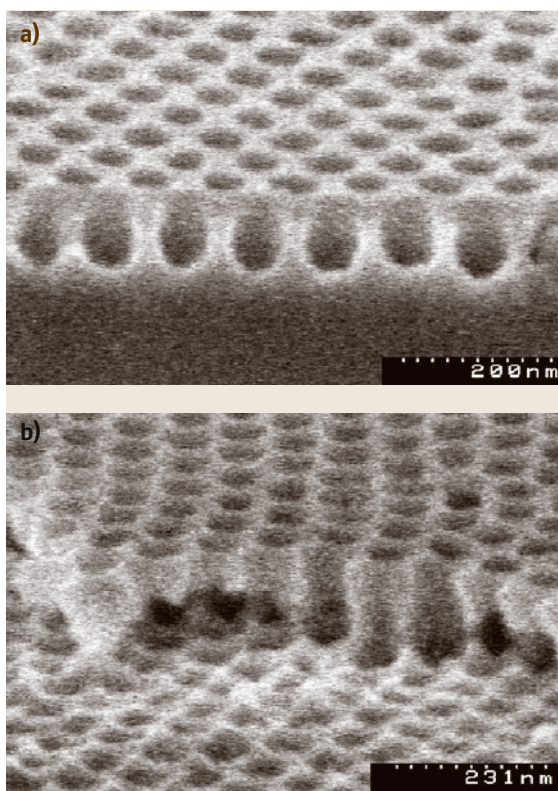


Fig. 41.6a,b Example of nanoscale features formed in a SrTiO₃ (100) substrate using conventional ion implantation through high-aspect-ratio masks combined with selective wet etching [41.51, 52]. Scanning electron microscope images for: (a) tilt view of single implant and (b) multiple energy implants to create membrane structure

as a substitute for lithographic pre-patterning for the fabrication of ordered nanochannels in materials such as Si and Al_2O_3 [41.46].

41.2.2 3-D PC Structures

Introduction to 3-D PBG Structures

A main advantage of 3-D PBG structures over 2-D PBG structures is that the 3-D PBG structures can overcome the problem of diffraction losses in the third (out-of-plane) dimension [41.1, 2]. However, this advantage is countered by the significant challenges associated with fabricating 3-D PBG structures. Fabricated 3-D structures reported to date lack the reproducibility and design versatility required for functional devices [41.21].

3-D Lithographic Fabrication Methods

Fabrication of 3-D PCs using top-down micromachining techniques has been more difficult than for 2-D PCs.

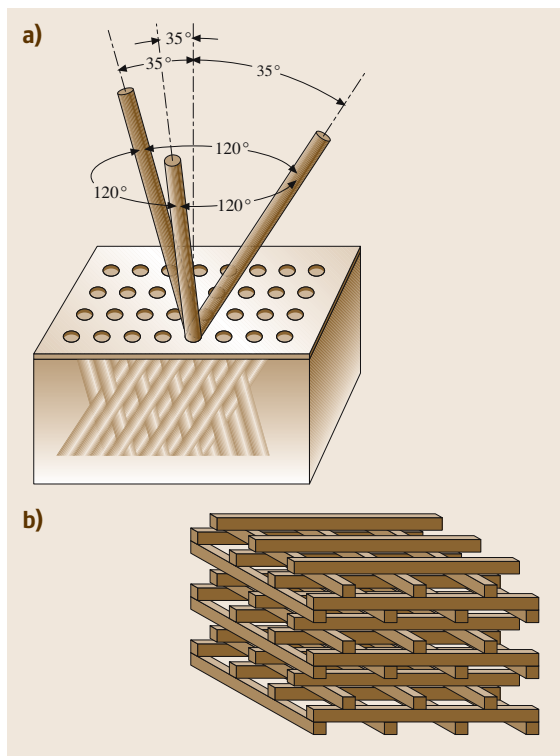


Fig. 41.7a,b Examples of 3-D photonic crystal structures: (a) Yablonovite, obtained by drilling holes in a homogeneous dielectric material (15 GHz); and (b) wood-pile structure, obtained by stacking micromachined Si rods (500 GHz) [41.2]

The first experimental demonstration of a full 3-D photonic band gap was performed in the microwave region (i.e., at 15 GHz, $\lambda \approx 2$ cm), due to the relative ease of fabrication of appropriate structures on the centimeter scale [41.58] (Fig. 41.7). Another case of a full PBG was demonstrated at 500 GHz ($\lambda \approx 600 \mu\text{m}$) with a 3-D structure fabricated by assembling micromachined silicon dielectric rods [41.59] (Fig. 41.7) into log-pile structures [41.60, 61].

In a 3-D lattice, theoretical calculations have determined that the band gap position is generally about one half to one third of the periodicity scale [41.62]. This translates into patterning the dielectric materials with features of 100–300 nm in all three dimensions, with required alignment accuracies among the different layers exceeding tens of nanometers [41.62]. In addition, with air as the second dielectric, the empty volume required for large PBGs is so large (≈ 70 –80%) that the mechanical stability of the structure becomes very poor, adding another challenge for the etching procedure [41.2]. Often 3-D PC structures are so full of defects after such etching that the systems no longer exhibit significant PC behavior [41.2].

Due to difficulties in fabricating structures in the optical or near-infrared frequencies [41.2, 62, 63], only a few experimental systems have recently demonstrated the signature of a complete PBG. Lithographically derived 3-D structures require many complex fabrication steps, and it takes many months to make a single structure. Some examples of prototype 3-D PC include log-pile structures fabricated using layer-by-layer patterning (i.e., repeated deposition and etching) of Si [41.64] (Fig. 41.8), and structures fabricated using wafer fusion combined with etching in GaAs and InP [41.65] (Fig. 41.8). In the former case, processing difficulties have limited the number of layers in such 3-D periodic structures (i.e., to below 10) and in the latter cases, to materials other than those commonly used in microelectronics [41.62]. Sequential deposition of dielectric films on a patterned substrate may be used to produce 3-D periodic structures in the *auto-cloning* process [41.66], which is considered to be a relatively cheap and flexible process. However, the range of structures that can be produced with this technique is limited. Also, the technique does not allow the component dielectrics to be continuously connected in 3-D [41.67].

Hole-drilling has also been implemented using X-ray lithography [41.68] and can be scaled to optical frequencies by directional etching through a mask on a semiconductor surface [41.69]. Other

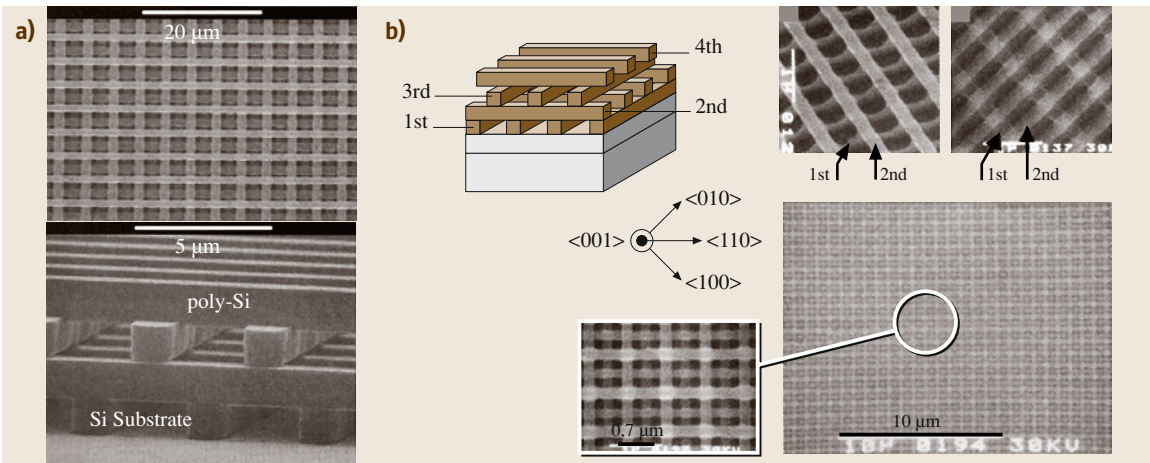


Fig. 41.8a,b Examples of 3-D PBG structures fabricated by micromachining/lithographic techniques: (a) advanced lithography [41.64], (b) wafer fusion [41.65]

fabrication methods include direct laser writing, but its resolution is currently only suitable for infrared structures.

3-D Self-Assembly Methods: Inverse Opals

The technological challenges involved in extending microlithographic techniques to 3-D patterning have made self-assembly methods attractive for 3-D PC fabrication. The benefits of using this type of bottom-up assembly for 3-D fabrication include its low capital cost, and its parallel and rapid nature, which allows the ordered fabrication of 3-D PBG-type structures over macroscopic length scales [41.3, 62, 73, 74]. However, self-assembly methods for 3-D PC fabrication are often material-specific, and templated bottom-up PCs tend to have diluted properties compared to bulk single crystals, typically containing unintentional defects and lacking reproducibility.

Self-organized 3-D colloidal crystal structures in a face-centered-cubic lattice (i.e., artificial opals) can be fabricated through sedimentation of colloidal suspensions (i.e., typically spheres, but other structures, such as micelles formed from rod-coil block copolymers have been used). Previously, colloidal crystals were not considered appropriate for PCs with band gaps in the visible and IR ranges because the colloids [typically SiO₂ or polymer materials such as polystyrene (PS) [41.75, 76]

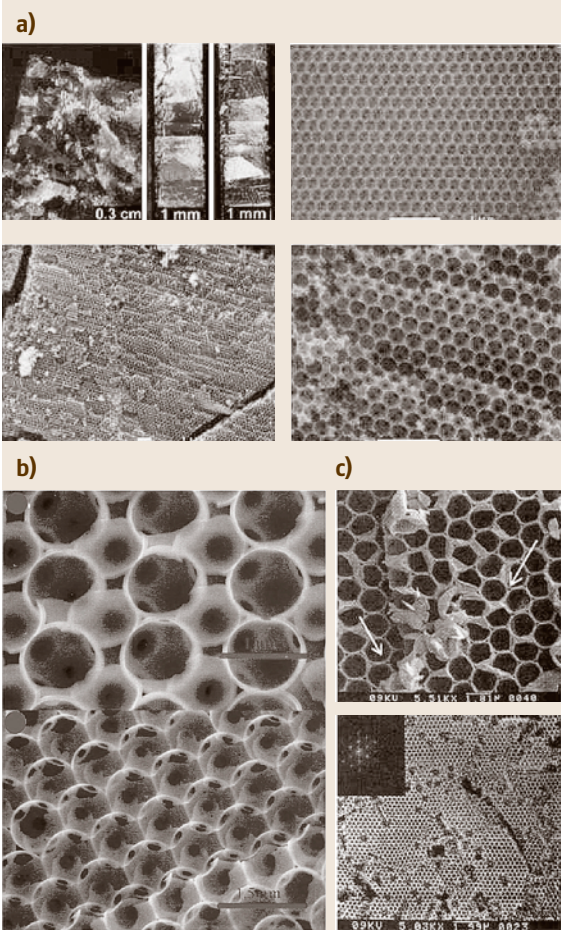


Fig. 41.9a–c Examples of 3-D inverse opal PBG structures: (a) graphite (using CVD infiltration) [41.70], (b) Si (using CVD infiltration) [41.71], (c) titania (using liquid-phase chemical reaction) [41.72]

or poly (methyl-methacrylate) (PMMA)] have refractive indices ≈ 1.5 , which are too low to obtain a full PBG [41.77–79]. Such structures have weak refractive-index contrast, but theoretical calculations have shown that the final porosity of the artificial opal is too small to ever obtain a full PBG, even with very high refractive-index contrast [41.1, 3, 73].

These colloidal structures may also be infiltrated with a second material of higher refractive index to potentially improve the dielectric contrast [41.77, 80], while providing the required dimensional control and low solid fraction [41.74]. So called *inverse* structures can be formed when the original colloidal structure is removed to leave a macroporous structure. Calculations have shown that these inverse structures give rise to larger PBGs for the same refractive-index contrast [41.81]. The porous materials obtained using this approach have also been referred to as *inverse opals* or *inverted opals* because they have an open, periodic 3-D framework complementary to that of the parent opal. Modeling suggests that a porous material consisting of an opaline lattice of interconnected air balls (embedded in an interconnected matrix with a higher refractive index) should produce a full PBG. Optimum photonic effects require that the volume fraction of the matrix material should be 20–30%. Template-based synthesis based on colloids is attractive because the periodicity of this system can be conveniently tuned and a wide variety of materials with relatively high refractive index can be used [41.73]. There have been many examples of inverse opals using materials such as carbon, titania, zirconia, Si and Ge reported in the literature [41.3, 70–74, 82–89] (Fig. 41.9).

The inverse structure is made by infiltrating the interstitial void spaces ($\approx 26\%$ by volume) between the colloidal spheres in the opaline array with a fluid precursor. The fluid may be a pure liquid (e.g., a liquid metal alkoxide or a molten metal), a solution (e.g., a salt solution), a vapor (e.g., using CVD), or a colloidal dispersion of nanocrystals (e.g., colloidal gold particles) [41.73, 74]. Alternately, suspensions containing nanoparticle precursors (with sizes of ≈ 1 –50 nm) and monodisperse spheres can be co-precipitated into an ordered structure [41.73, 74]. The void spaces between the colloidal spheres has also been filled using a variety of materials by electrochemical deposition [41.73]. The fidelity of this procedure is mainly determined by van der Waals interactions, the wetting of the template surface, kinetic factors such as the filling of the void spaces in the template, and the volume shrinkage of precursors during solidification [41.73].

The choice of wall material depends on the desired optical functionality of the PC, as its properties depend on the transparency region of the dielectric material. The availability of a precursor that can infiltrate the voids between the colloidal spheres without significantly swelling or dissolving the template is critical. For example, high-temperature infiltration techniques exclude the use of polymer templates. The precursor must be carefully chosen to allow for shrinkage occurring during solidification and the final grain size, which affects the smoothness, density, effective refractive index, PC wall structure and the mechanical properties of the final 3-D PC [41.74]. The approach and conditions used for infiltration can affect the filling efficiencies (i.e., the effective wall density) which, in turn, affect the mechanical strength and reduces the average refractive index of the walls [41.74, 82]. Choice of template also depends on the method of solidification of the precursor and whether the template can be selectively removed from the wall material, either chemically or by heating during the final fabrication step.

Porous structures are produced by the removal of the colloidal templates, either by burning, extracting with a solvent, or, in the case of silica colloids, by dissolving with dilute hydrofluoric (HF) acid. The method of template removal depends on the colloid material and on the properties of the infiltrated precursor. Silica sphere template arrays are removed by dissolution in aqueous HF solutions [41.74, 88]. Most metal-oxide inverse opals have typically been synthesized using polymer spheres as templates [41.74]. Polymer sphere template arrays are often removed thermally, with the added benefit of chemically converting the sol-gel precursors in the infiltrated phase to the desired oxide product [41.74, 88]. If precursor solidification is feasible at low temperatures, or if the precursor has components that would be destroyed by thermal template removal, polymer spheres can also be removed chemically using organic solvents, such as toluene, or tetrahydrofuran (THF) and acetone mixtures, or by photodegradation [41.74, 88].

Fabrication of such structures is not trivial: some of the difficulties include incomplete infiltration, structural shrinkage after infiltration, and microstructural variations (affecting the refractive index). A major problem with chemical assembly is that residual disorder drastically reduces the PBG width and the intensity of the reflectance peaks [41.74, 85]. Such disorder can arise from several different sources. Firstly, there are defects in the initial opal template, such as stacking faults or small deviations in sphere size or shape. It is not clear whether the imperfections present in current in-

verse opals are small enough to observe a complete PBG [41.3]. While progress has been made in reducing such defects, numerical simulations predict that even a 2% deviation in sphere size (or lattice constant) can close a PBG in even high-refractive-index inverted opals [41.3, 74]. Secondly, significant disorder can arise from the infiltration process itself. Even the best samples to date have variations in the amount of guest material filling the lattice [41.3]. Since these inhomogeneities must be much smaller than the optical wavelength, the elimination of residual order remains a challenge with such methods [41.3].

Although 3-D inverse opals are fascinating research objects, they are not amenable for practical photonic applications because of difficulties in integrating them into PLCs. Thus, there is strong interest in fabricating 3-D inverse opals in pseudo-2-D form, or supported on a substrate as an inverse opal film, to be compatible with traditional optoelectronic elements for system integration [41.1, 86, 91–93]. Recently, ordered colloidal crystal films have been formed by convective assembly [41.86] by particle confinement in thin slits between two solid plate boundaries [41.73, 86] and by compacting particles in the dry state [41.86, 91, 93–96]. Recently, patterned substrates were successfully used to control slow sedimentation of colloidal particles to improve the long-range order and to induce growth along a specific crystal plane [41.74, 86]. However, growth of defect-free crystalline templates with desired symmetry and orientation remains one of the challenges of the field.

Liquid sol–gel precursors can be used to form a variety of oxide materials and are very convenient for infiltration into artificial opal templates [41.88, 97]. However, the typical high percentage of volume shrinkage (typically 15–30%) involved in the solidification of such sol–gel precursors during the precursor-to-oxide conversion process has often resulted in random cracks in the final inverse opal films [41.73, 74, 88], often preventing the fabrication of large, ordered, inverse opal film structures. Colloidal template films have been infiltrated by soaking the template into the precursor sol–gel solution [41.74, 85, 98], but achieving a dense, complete filling of the void spaces within the 3-D template film has proven to be challenging [41.97]. Another approach used for large-area inverse opal film fabrication involves the crystallization of colloids in a liquid medium containing nanoparticles of the desired material, which are subsequently packed into a dense, 3-D lattice [41.74, 97]. The main limitation of this method is that nanometer-sized particles are not always available for the desired wall ma-

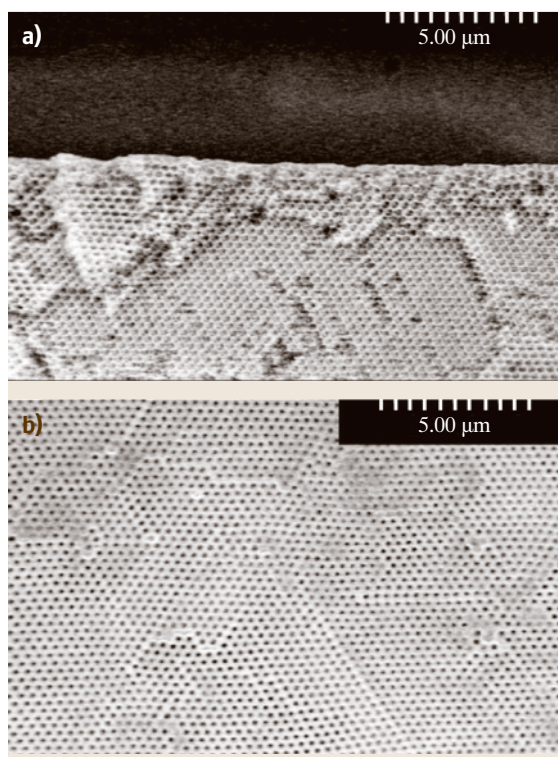


Fig. 41.10a,b Scanning electron microscope images [41.90] of Pb-doped BST inverse opal thin film: (a) cross section; and (b) top view

terial, and so only very specific materials may be made into inverse opals [41.88]. Another method of fabricating high-quality inverse opal films derived from liquid precursors is by infiltrating a sol–gel liquid precursor into a colloidal crystal template film by spin-coating, which results in very uniform filling of the template [41.90] (Fig. 41.10). By comparison, solvent-extracted samples shrink about 10% less [41.88]. Thus, for organic polymers, it has been possible to fabricate such 3-D porous structures as free-standing thin films that are several square centimeters in area.

Alternative 3-D PBG Fabrication Methods

Alternative approaches to lithography or self-assembly for 3-D fabrication exist. Some notable examples include laser-guided stereolithography, in which structures can also be written directly layer-by-layer by using a tightly focused laser beam to cause chemical vapor deposition [41.99], and two-photon-induced deposition or polymerization [41.100, 101]. Currently, feature sizes that can be routinely achieved using these methods are

on the order of a few micrometers due to focusing and diffraction effects.

A 3-D holographic patterning process using multiple laser beams has recently been investigated; a 3-D laser interference pattern exposes a photosensitive polymer precursor (i. e., a photoresist) rendering the exposed areas insoluble and unexposed areas are dissolved away

to leave a 3-D PC formed of cross-linked polymer with air-filled voids [41.102]. This process allows flexible design of the structure of a unit cell and thus of the optical properties of the microstructured material. Of concern is that such structures have relatively weak refractive-index contrast, and the processing approach is quite demanding [41.62].

41.3 Tunable PCs

Although conventional PCs offer an ability to control light propagation or confinement through the introduction of defects, once such defects are introduced, the propagation or confinement of light in these structures is not controllable. Thus, discretionary switching of light, for example, or rerouting of optical signals, is not available with fixed defects in PCs. There are two approaches that have been pursued to tune the properties of PCs: these are by tuning the refractive index of the constituent materials, or by altering the physical structure of the PC. In the latter case, the emphasis has been principally on changing the lattice constant, although other approaches are also relevant (i. e., including for example, the fill factor, structure symmetry and scattering element shape). In the text below, we discuss the state of the art in both of these areas.

41.3.1 Tuning the PC Response by Changing the Refractive Index of the Constituent Materials

We discuss recent progress in using four approaches within this category – namely, tuning the PC response by using: (i) light, (ii) applied electric fields, (iii) temperature or electrical field in infiltrated PCs, and (iv) by changing the concentration of free carriers (using electric field or temperature) in semiconductor-based PCs.

PC Refractive-Index Tuning Using Light

One approach to modifying the behavior of PCs is to use intense illumination of the PC by one beam of light to change the optical properties of the crystal in a nonlinear fashion; this, in turn, can thus control the properties of the PC for another beam of light. An example of a nonlinear effect includes the flattening of the photon dispersion relation near a PBG, which relaxes the constraints on phase matching for second- and third-harmonic generation [41.103]. In addition, light location near defects can enhance a variety of third-order nonlinear processes in 1-D PCs [41.104].

A new approach for PBG tuning has been proposed based on photo-reversible control over molecular aggregation, based on using the photochromic effect in dyes [41.105]. This can cause a reversible change over the photonic stop band. Structures that were studied included opal films formed from 275-nm-diameter silica spheres, infiltrated with photochromic dye: two dyes were considered – namely, 1,3-dihydro-1,3,3-trimethylspiro-[2H-indol-2,3'-[3H]-naphth[2,1-b] [1,4]oxazine] (SP) and cis-1,2-dicyano-1,2-bis(2,4,5-trimethyl-3-trienyl)ethane (CMTE). For the SP dye opal, a reversible 15-nm shift in the reflectance spectrum was observed following UV irradiation, which was ascribed to changes in the reflective index due to resonant absorption near the stop band. Smaller shifts (of about 3 nm) were observed for the CMTE dye opal. The recovery process on cessation of UV illumination was quite slow, taking about 38 s for the SP dye.

Nonlinear changes in refractive index have also been studied in PCs consisting of 220-nm-diameter SA (self assembly) polystyrene spheres infiltrated with water [41.106]. In these studies, the optical Kerr effect was used to shift the PBG. 40 GW/cm² [41.2] of peak pump power at 1.06 μ m (35-ps pulses at 10-Hz repetition rate) was used to shift the PBG 13 nm. The large optical nonlinearity originates from the delocalization of conjugated π -electrons along the polymer chains, leading to a large third-order nonlinear optical susceptibility. The time response was measured as a function of the delay time between pump and probe, and confirmed a response time of several picoseconds.

PC Refractive-Index Tuning Using an Applied Electric Field

There have been a number of studies focusing on using ferroelectric materials to form PCs [41.90, 98, 107, 108]. The application of an electric field to such materials can be used to change the refractive index and tune the PC optical response. For example, lead lanthanum zirconate titanate (PLZT) inverse opal structures have been fabri-

cated by infiltration using 350-nm-diameter polystyrene sphere templates and annealing at 750 °C [41.107]. The films were formed on indium tin oxide (ITO) coated glass to enable electric-field-induced changes in reflectivity to be measured due to the electro-optic effect. Applying voltages of up to about 700 V across films of thickness of about 50 nm only achieved a few nm of peak shift, attributed to the very modest changes in refractive index from the applied field (i.e., from 2.405 to 2.435, as a result of the bias). It should be noted, however, that the intrinsic response of the electro-optic effect in these materials is known to be in the GHz range and hence highly suitable for rapid tuning. Other reports include the formation of inverse opal barium strontium titanate (BST) PCs using infiltration of polystyrene opals [41.90, 98]. BST is most interesting in that it provides a high-refractive-index material and a factor of at least two times higher breakdown field strength than PLZT, and hence offers a much wider range of applied fields for tuning. Reports of using other ferroelectric materials include a high-temperature infiltration process for the ferroelectric copolymer, poly(vinylidene fluoride-trifluoroethane), infiltrated into 3-D silica opals with sphere diameters of 180, 225 and 300 nm [41.108].

Refractive-Index Tuning of Infiltrated PCs

In this case the approach has been to consider modulation of the refractive index of a PC infiltrated with a tunable medium – in particular, the most popular approach has been to use liquid crystals to infiltrate porous 2-D and 3-D PC structures. Such liquid crystals can behave as ferroelectrics whose refractive index may be tuned using either an applied electric field, or by thermal tuning. Reported results for this approach, using 3-D inverse opal structures, have been restricted to infiltration of ferroelectric liquid-crystal material into a silicon/air PC [41.109]. Reported changes in the refractive index of the liquid crystal are 1.4–1.6 under an applied field [41.109]. However, since the ferroelectric liquid has a higher refractive index than air, the infiltration results in a significant decrease in the refractive index contrast. This means that the original full photonic band gap of the inverse opal silicon PC no longer exists, and the practical utility of the silicon structure as a PC is effectively lost. Theoretical simulations have shown that partial surface wetting of the internal inverse opal surface can retain the full photonic band in silicon [41.109], but it is questionable whether such a complex structure can ever be practically fabricated. Finally, the presence of ferroelectric liquid crystal surrenders one of the main advantages of the original concept for PC structures: that

is, permitting light propagation in air [41.110]. Temperature tuning of the liquid-crystal material was shown to result in very small changes in refractive index (changes in $n < 0.01$ over a 70 °C change) and thus only provide minimal shifts in the transmittance through the PC over a large temperature range [41.111].

PC Refractive-Index Tuning by Altering the Concentration of Free Carriers (Using Electric Field or Temperature) in Semiconductor-Based PCs

An elegant way to rapidly tune the PBG of semiconductor-based PCs is to adjust the refractive index by modulating the free-carrier concentration using an ultra-fast optical pulse [41.112]. Using this approach, reflectivity of a two-dimensional silicon-based honeycomb PCs with 412-nm air holes (100 μm in length) in a 500-nm periodic array, was studied with a pump–probe approach [41.112]. By varying the delay between the pump the speed of PBG tuning was measured to be about 0.5 ps. The reflectance relaxation (corresponding to the return of the PBG to its original position) occurred on a timescale of 10–100 ns, characteristic of recombination of excess electrons and holes. Although these results are very encouraging, this approach cannot suppress or correct for light-scattering losses caused by structural imperfections, which remain an important consideration for currently fabricated PCs.

4.1.3.2 Tuning PC Response by Altering the Physical Structure of the PC

The second approach that we discuss for tuning the response of a PC is based on changes to the physical structure of the PC. We discuss the following approaches for tuning using this approach: tuning using (i) temperature, (ii) an applied magnetic field, (iii) strain/deformation, (iv) piezoelectric effects, and (v) using micro-electro-mechanical systems (MEMS) [i.e., actuation].

Tuning PC Response Using Temperature

An example of this approach is a study of temperature tuning of PCs fabricated from self-assembled polystyrene beads [41.113]. The PBG in these structures was fine-tuned by annealing samples at temperatures of 20–100 °C, resulting in a continuous blue shift of the stop band wavelength from 576 nm to 548 nm. New stop bands appeared in the UV transmission spectra when the sample was annealed above about 93 °C – the glass-transition temperature of the polystyrene beads.

Tuning PC Response Using Magnetism

An example of this approach is the use of an external applied magnetic field to adjust the spatial orientation of a PC [41.97]. This can find application in fabricating photonic devices such as tunable mirrors and diffractive display devices. These authors fabricated magnetic PCs by using monodisperse polystyrene beads self-assembled into a ferro-fluid consisting of magnetite particles, with particle sizes below 15 nm [41.97]. On evaporation of the solvent, a cubic PC lattice was formed with the nanoparticles precipitating out into the interstices between the spherical polystyrene colloids. These authors then showed how the template could be selectively removed by calcination or wet etching to reveal an inverse opal of magnetite – such structures being proposed as suitable for developing magnetically tunable PCs.

Tuning PC Response Using Strain

The concept in this case is quite straightforward – deforming or straining the PC changes the lattice constant or arrangement of dielectric elements in the PC with concomitant change in the photonic band structure. Polymeric materials would appear to be most suited to this methodology, owing to their ability to sustain considerable strains. However, concerns of reversibility and speed of tuning would clearly need to be addressed. Theoretical predictions for the influence of deformation on such systems include a report on a new class of PC based on self-assembling cholesteric elastomers [41.114]. These elastomers are highly deformable when subjected to external stress. The high sensitivity of the photonic band structure to strain, and the opening of new PBGs have been discussed [41.114]. Charged colloidal crystals were also fixed in a poly(acrylamide) hydrogel matrix to fabricate PCs whose diffraction peaks were tuned by applying mechanical stress [41.115]. The PBG shifted linearly and reversibly over almost the entire visible spectral region (from 460 nm to 810 nm).

Modeling of the photonic band structure of 2-D silicon-based triangular PCs under mechanical deformation was also reported [41.116]. The structures considered consisted of a silicon matrix with air columns. The authors showed that while a 3% applied shear strain provides only minor modifications to the PBG, uniaxial tension can produce a considerable shift. Other modeling includes a study of how strain can be used to tune the anisotropic optical response of 2-D PCs in the long-wavelength limit [41.117]. These calculations showed that the decrease in dielectric constant per unit strain is larger in the direction of the strain

than normal to it. Indeed, the calculated birefringence is larger than that of quartz. They suggest that strain-tuning of this birefringence has attractive application in polarization-based optical devices.

To appreciate the sensitivity of such structures to mechanical tuning, it is instructive to refer to some recent work on PMMA inverse opal PC structures that were fabricated using silica opal templates [41.118]. Under the application of uniaxial deformation of these PCs, the authors found a blue shift of the stop band in the transmission spectrum – the peak wavelength of the stop band shifted from about 545 nm in the undeformed material to about 470 nm under a stretch ratio of about 1.6.

Another practical approach that has been applied to physically tuning PC structures is that of thermal annealing. One such study showed how the optical properties of colloidal PCs consisting of silica spheres can be tuned through thermal treatment [41.119]. This was attributed to both structural and physio-chemical modification of the material on annealing. A shift in the minimum transmission from about 1000 nm (un-annealed) to about 850 nm (for annealing at about 1000 °C) was demonstrated, or about a maximum shift in Bragg wavelength of $\approx 11\%$.

A quite novel application of strain-tuning was recently reported [41.120]. The authors studied 2-D PCs consisting of arrays of coupled optical microcavities fabricated from vertical-cavity surface-emitting laser structures. The influence of strain, as manifested by shifts in the positions of neighboring rows of microcavities with respect to each other, corresponded to alternating square or quasi-hexagonal shear-strain patterns. For strains below a critical threshold value, the lasing photon mode-locked to the corresponding mode in the unstrained PC. At the critical strain, switching occurred between the square and hexagonal lattice modes.

Finally, there has been a proposal for using strain in a PC to tune the splitting of a degenerate photon state within the PBG, suitable for implementing tunable PC circuits [41.121]. The principle applied is analogous to the static Jahn–Teller effect in solids. These authors showed that this effect is tunable by using the symmetry and magnitude of the lattice distortion. Using this effect the design of an optical valve that controls the resonant coupling of photon modes at the corner of a T-junction waveguide structure has been discussed.

Tuning PC Response Using Piezoelectric Effects

In this section we discuss using piezoelectric effects to physically change PC structures and hence tune them.

A proposal was made for using the piezoelectric effect to distort the original symmetry of a two-dimensional PC from a regular hexagonal lattice to a quasi-hexagonal lattice under an applied electric field [41.122]. The original bands decomposed into several strained bands, dependent on the magnitude and direction of the applied field. In the proposed structures, the application of $\approx 3\%$ shear strain is shown to be suitable for shifting 73% of the original PBG, which they refer to as the tunable-band-gap regime. An advantage of such an approach is that such structures are suggested to be capable of operation at speeds approaching MHz. Another report [41.123] discusses the design and implementation of a tunable silicon-based PBG microcavity in an optical waveguide, where tuning is accomplished using the piezoelectric effect to strain the PC; this was carried out using integrated piezoelectric microactuators. These authors report on a 1.54-nm shift in the cavity resonance at 1.56 μm for an applied strain of 0.04%.

There have also been reports of coupling piezoelectric-based actuators to PCs. One such report [41.124] discusses a poly(2-methoxyethyl acrylate)-based PC composite directly coupled to a piezoelectric actuator to study static and dynamic stop-band tuning characteristics; the stop band of this device could be tuned through a 172-nm tuning range, and could be modulated at up to 200 Hz.

Tuning PC Response Using Micro-Electro-Mechanical Systems (MEMS) Actuation

There have been a number of interesting developments in this field including PC-based air-bridge devices consisting of suspended 1-D PC mirrors separated by a Fabry–Perot cavity (gap) [41.125]. When such structures are mechanically perturbed, there can be a substantial shift in the PBG due to strain. The au-

thors discuss how a suite of spectrally tunable devices can be envisioned based on such structures – these include modulators, optical filters, optical switches, WDM (wavelength division multiplexing), optical logical circuits, variable attenuators, power splitters and isolators. Indeed, the generalization of these concepts beyond 1-D was discussed in a recent patent [41.126] that covers tunable PC structures. This report [41.126], as well as others [41.127], consider the extension of these ideas to form families of micromachined devices. These authors [41.127] modeled and implemented a set of micromachined vertical resonator structures for 1.55- μm filters consisting of two PC (distributed Bragg reflector [DBR]) mirrors separated by either an air gap or semiconductor heterostructure. Electro-mechanical tuning was used to adjust the separation between the mirrors and hence fine-tune the transmission spectrum. The mirror structures were implemented using strong-index-contrast InP/air DBRs giving an index contrast of 2.17, and weak-index-contrast (0.5) silicon nitride/silicon dioxide DBRs. In the former case, a tuning range of over 8% of the absolute wavelength was achieved; varying the inter-membrane voltage up to 5 V gave a tuning range of over 110 nm. Similar planar structures are discussed in other papers [41.128] where the mirrors are formed from two slabs of PC separated by an adjustable air gap [41.126]. These structures have been shown to be able to perform as either flat-top reflection or all-pass transmission filters, by varying the distance between the slabs, for normally incident light. Unlike all previously reported all-pass reflection filters, based on Gires–Tournois interferometers using multiple dielectric stacks, their structure generates an all-pass transmission spectrum, significantly simplifying signal extraction and optical alignment – also the spectral response is polarization-independent owing to the 90° rotational symmetry of their structure.

41.4 Summary and Conclusions

This chapter discussed the fabrication and properties of PCs, with emphasis on developing tunable structures. New fabrication methodologies and also tuning schemes offers a glimpse of the far-reaching prospects for developing photonic devices which can, in a discretionary fashion, control the propagation of modes of light in an

analogous fashion to the way in which nanostructures have been harnessed to control electron-based phenomena. Analogous to the evolution of electronic systems, one can anticipate a path toward development of compact active integrated photonic systems, as envisioned based on the technology outlined in this review.

References

- 41.1 T. F. Krauss, R. M. de la Rue: Photonic crystals in the optical regime – past, present, and future, *Prog. Quantum Electron.* **23**, 51–96 (1999)
- 41.2 V. Berger: From photonic band gaps to refractive index engineering, *Opt. Mater.* **11**, 131–142 (1999)
- 41.3 D. J. Norris, Y. A. Vlasov: Chemical approaches to three-dimensional semiconductor photonic crystals, *Adv. Mater.* **13**(6), 371–376 (2001)
- 41.4 V. Mizeikis, S. Juodkazis, A. Marcinkevicius, S. Matsuo, H. Misawa: Tailoring and characterization of photonic crystals, *J. Photochem. Photobiol. C* **2**, 35–69 (2001)
- 41.5 Photonics Nanostruct. **1**(1), 1–78 (2003): whole edition is dedicated to fundamentals and applications of photonic crystals
- 41.6 S. G. Johnson, J. D. Joannopoulos: *Photonic Crystals: Road from Theory to Practice* (Kluwer Academic, Boston 2002)
- 41.7 J. D. Joannopoulos, R. D. Meade, J. N. Winn: *PCs: Moulding the Flow of Light* (Princeton Univ. Press, Princeton 1995)
- 41.8 S. John: Strong localization in certain disordered dielectric super-lattices, *Phys. Rev. Lett.* **58**(23), 2486–2489 (1987)
- 41.9 E. Yablonovitch: Inhibited spontaneous emission in solid state physics and electronics, *Phys. Rev. Lett.* **58**(20), 2059–2062 (1987)
- 41.10 H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sato, S. Kwakami: Superprism phenomena in photonic crystals, *Phys. Rev. B* **58**(16), R10096–R10099 (1998)
- 41.11 V. Berger: Photonic crystals and photonic structures, *Cur. Opin. Solid State Mater. Sci.* **4**, 209–216 (1999)
- 41.12 K. M. Ho, C. T. Chan, C. M. Soukoulis: Existence of a photonic gap in periodic dielectric structures, *Phys. Rev. Lett.* **65**(25), 3152–3155 (1990)
- 41.13 S. Satpathy, Z. Zhang, M. R. Salehpour: Theory of photon bands in three-dimensional periodic dielectric structures, *Phys. Rev. Lett.* **64**(11), 1239–1242 (1990)
- 41.14 E. Yablonovitch: Photonic band-gap structures, *J. Opt. Soc. Am. B* **10**(2), 283–295 (1993)
- 41.15 Y. Xia: Photonic Crystals, *Adv. Mater.* **13**(6), 369 (2001)
- 41.16 C. Anderson, K. Giapis: Larger two-dimensional photonic band gaps, *Phys. Rev. Lett.* **77**, 2949–2952 (1996)
- 41.17 R. D. Meade, A. M. Rappe, K. D. Brommer, J. D. Joannopoulos: Nature of the photonic band gap: Some insights from a field analysis, *J. Opt. Soc. Am. B* **10**(2), 328–332 (1993)
- 41.18 R. D. Meade, A. M. Rappe, K. D. Brommer, J. D. Joannopoulos, O. L. Alerhand: Accurate theoretical analysis of photonic band-gap materials, *Phys. Rev. B* **48**(11), 8434–8437 (1993)
- 41.19 C. Lopez: Materials aspects of PCs, *Adv. Mater.* **15**(20), 1679–1704 (2003)
- 41.20 C. Weisbuch, C. H. Benisty, S. Olivier, M. Rattier, C. J. M. Smith, T. F. Krauss: Advances in photonic crystals, *Phys. Status Solidi B* **221**(93), 93–99 (2000)
- 41.21 C. Weisbuch, H. Benisty, M. Rattier, C. J. M. Smith, T. F. Krauss: Advances in 2D semiconductor PCs, *Synth. Mater.* **116**, 449–452 (2001)
- 41.22 S. L. Swartz: Topics in electronic ceramics, *IEEE Trans. Electr. Ins.* **25**(5), 935–987 (1990)
- 41.23 F. Watt: Focused high energy proton beam micromachining: A perspective view, *Nucl. Instrum. Methods Phys. Res.* **158**, 165–172 (1999)
- 41.24 D. K. Ferry, R. O. Grondin: *Physics of Submicron Devices* (Plenum, New York 1991)
- 41.25 J. Gierak, D. Mailly, G. Faini, J. L. Pelouard, P. Denk, F. Pardo, J. Y. Marzin, A. Septier, G. Schmid, J. Ferre, R. Hydman, C. Chappert, J. Flicstein, B. Gayral, J. M. Gerard: Nano-fabrication with focused ion beams, *Microelectron. Eng.* **57–58**, 865–875 (2001)
- 41.26 K. Gamo: Nanofabrication by FIB, *Microelectron. Eng.* **32**, 159–171 (1996)
- 41.27 J. Melngailis, A. A. Mondelli, I. L. Berry III, R. Mohondro: A review of ion projection lithography, *J. Vac. Sci. Technol. B* **16**(3), 927–957 (1998)
- 41.28 P. Peercy: The drive to miniaturization, *Nature* **406**, 1023–1026 (2000)
- 41.29 T. Ito, S. Okazaki: Pushing the limits of lithography, *Nature* **406**, 1027–1031 (2000)
- 41.30 N. Peng, C. Jeynes, R. P. Webb, I. R. Chakarov, M. G. Blamire: Monte Carlo simulations of masked ion beam irradiation damage profiles in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin films, *Nucl. Instrum. Methods Phys. Res. B* **178**, 242–246 (2001)
- 41.31 J. L. Bartelt: Masked ion beam lithography: An emerging technology, *Sol. State Technol.* **29**(5), 215–220 (1986)
- 41.32 D. P. Stumbo, J. C. Wolfe: Contrast of ion beam proximity printing with non-ideal masks, *J. Vac. Sci. Technol. B* **12**(6), 3539–3542 (1994)
- 41.33 T. Devolder, C. Chappert, Y. Chen, E. Cambril, H. Launois, H. Bernas, J. Ferre, J. P. Jamet: Patterning of planar magnetic nanostructures by ion irradiation, *J. Vac. Sci. Technol. B* **17**(6), 3177–3181 (1999)
- 41.34 P. Ruchhoeft, J. C. Wolfe, R. Bass: Ion beam aperture-array lithography, *J. Vac. Sci. Technol. B* **19**(6), 2529–2532 (2001)
- 41.35 Y. Hsieh, Y. Hwang, J. Fu, Y. Tsou, Y. Peng, L. Chen: Dislocation multiplication inside contact holes, *Microelectron. Reliab.* **39**, 15–22 (1999)

- 41.36 J. Gierak, A. Septier, C. Vieu: Design and realization of a very high-resolution FIB nanofabrication instrument, *Nucl. Instrum. Methods Phys. Res. A* **427**, 91–98 (1999)
- 41.37 A. N. Broers, A. C. F. Hoole, J. M. Ryan: Electron beam lithography – resolution limits, *Microelectron. Eng.* **32**, 131–142 (1996)
- 41.38 C. Lehrer, L. Frey, S. Petersen, H. Ryssel: Limitations of focused ion beam nano-machining, *J. Vac. Sci. Technol. B* **19**(6), 2533–2538 (2001)
- 41.39 H. Masuda, K. Fukuda: Ordered metal nanohole arrays made by a two-step replication of honeycomb structures of anodic alumina, *Science* **268**, 1466–1468 (1995)
- 41.40 R. Tonucci, B. Justus, A. Campillo, C. Ford: Nanochannel array glass, *Science* **258**, 783–785 (1992)
- 41.41 V. Lehmann, H. Foll: Formation mechanism and properties of electrochemically etched trenches in n-type silicon, *J. Electrochem. Soc.* **137**(2), 653–659 (1990)
- 41.42 A. Birner, R. B. Wehrspohn, U. M. Gosele, K. Busch: Silicon-based photonic crystals, *Adv. Mater.* **13**(6), 377–388 (2001)
- 41.43 J. Martin: Nanomaterials: A membrane-based synthetic approach, *Science* **266**, 1961–1966 (1994)
- 41.44 J. I. Martin, J. Nogues, K. Liu, J. L. Vicent, I. K. Schuller: Ordered magnetic nanostructures: Fabrication and properties, *J. Magn. Mater.* **256**(1–3), 449–501 (2003)
- 41.45 H. Masuda, M. Ohya, H. Asoh, M. Nakao, M. Nohtomi, T. Tamamura: Photonic crystals using anodic porous alumina, *Jpn. J. Appl. Phys. Pt. 2* **38**(12A), L1403–L1405 (1999)
- 41.46 R. Wehrspohn, J. Schilling: Electrochemically prepared pore arrays for photonic-crystal applications, *MRS Bull.*, 623–626 (2001)
- 41.47 A. P. Li, F. Muller, A. B. K. Nielsch, U. Gosele: Hexagonal pore arrays with a 50–420 nm interpore distance formed by self-organization in anodic alumina, *J. Appl. Phys.* **84**(11), 6023–6026 (1998)
- 41.48 H. Masuda, H. Asoh, M. Watanabe, K. Nishio, M. Nakao, T. Tamamura: Square and triangular nanohole array architectures in anodic alumina, *Adv. Mater.* **13**, 189–192 (2001)
- 41.49 M. Nakao, S. Oku, T. Tamamura, K. Yasui, H. Masuda: GaAs and InP nanohole arrays fabricated by reactive beam etching using highly ordered alumina membranes, *Jpn. J. Appl. Phys. Pt. 1* **38**(2B), 1052–1055 (1999)
- 41.50 J. Liang, H. Chik, A. Yin, J. Xu: Two-dimensional lateral superlattices of nanostructures: Nonlithographic formation by anodic membrane template, *J. Appl. Phys.* **91**(4), 2544–2546 (2002)
- 41.51 N. Matsuura, T. W. Simpson, C. P. McNorgan, I. V. Mitchell, X. Mei, P. Morales, H. E. Ruda: Nanometer-scale pattern transfer using ion implantation. In: *Three-Dimensional Nano-engineered Assemblies*, MRS Proc., Vol. 739, ed. by T. M. Orlando, L. Merhari, D. P. Taylor, K. Ikuta (Mater. Res. Soc., Boston 2002)
- 41.52 N. Matsuura, T. W. Simpson, I. V. Mitchell, X. Mei, P. Morales, H. E. Ruda: Ultra-high density, non-lithographic, sub-100 nm pattern transfer by ion implantation and selective chemical etching, *Appl. Phys. Lett.* **81**(25), 4826–4828 (2002)
- 41.53 E. Rimini: *Ion Implantation: Basics to Device Fabrication* (Kluwer Academic, Norwell 1995)
- 41.54 G. Hobler: Monte Carlo simulation of two-dimensional implanted dopant distributions at mask edges, *Nucl. Instrum. Methods Phys. Res. B* **96**, 155–162 (1995)
- 41.55 M. M. Faye, C. Vieu, G. B. Assayag, P. Salles, A. Claverie: Lateral damage extension during masked ion implantation into GaAs, *J. Appl. Phys.* **80**(8), 4303–4307 (1996)
- 41.56 P. Schmuki, L. Erickson: Direct micro-patterning of Si and GaAs using electrochemical development of focused ion beam implants, *Appl. Phys. Lett.* **73**, 2600–2602 (1998)
- 41.57 K. Wang, A. Chelnokov, S. Rowson, P. Garouche, J. – M. Lourtioz: Three-dimensional Yablonovite-like photonic crystals by focused ion beam etching of macroporous silicon, *Mater. Res. Soc. Symp. Proc.* **637**, E1.4.1–E1.4.5 (2001)
- 41.58 E. Yablonovitch, T. Gmitter, K. Leung: Photonic band structure: The face-centered-cubic case employing non-spherical atoms, *Phys. Rev. Lett.* **67**, 2295–2298 (1991)
- 41.59 E. Ozbay, A. Abeyta, G. Tuttle, M. Tringides, R. Biswas, C. Chan, C. Soukoulis, K. Ho: Measurement of a three-dimensional photonic band gap in a crystal structure made of dielectric rods, *Phys. Rev. B* **50**, 1945–1948 (1994)
- 41.60 K. M. Ho, C. T. Chan, C. M. Soukoulis, R. Biswas, M. Sigalas: Photonic band gaps in three dimensions: New layer-by-layer periodic structures, *Sol. State Commun.* **89**(5), 413–481 (1994)
- 41.61 H. S. Sözüer, J. P. Dowling: Photonic band calculations for woodpile structures, *J. Mod. Opt.* **41**(2), 231–239 (1994)
- 41.62 Y. Xia, B. Gates, Z.-Y. Li: Self-assembly approaches to three-dimensional photonic crystals, *Adv. Mater.* **13**(6), 409–413 (2001)
- 41.63 A. Moroz: Three-Dimensional complete photonic-band-gap structures in the visible range, *Phys. Rev. Lett.* **83**(25), 5274–5277 (1999)
- 41.64 S. Lin, J. Fleming, D. Hetherington, B. Smith, R. Biswas, K. Ho, M. Sigalas, W. Zubrzycki, S. Kurtz, J. Bur: A three-dimensional photonic crystals operating at infrared wavelengths, *Nature* **394**, 251–253 (1998)
- 41.65 S. Noda, K. Tomoda, N. Yamamoto, A. Chutinan: Full three-dimensional photonic bandgap crystals at near-infrared wavelengths, *Science* **289**, 604–606 (2000)

- 41.66 S. Kawakami: Fabrication of sub-micrometre 3D periodic structures composed of Si/SiO₂, *Electron. Lett.* **33**(4), 1260–1261 (1997)
- 41.67 C. T. Chan, S. Datta, K. M. Ho, C. M. Soukoulis: A7 structure: A family of photonic crystals, *Phys. Rev. B* **50**(3), 1988–1991 (1994)
- 41.68 G. Feiertag, W. Ehrfeld, H. Freimuth, H. Kolle, H. Lehr, M. Schmidt, M. M. Sigalas, C. M. Soukoulis, G. Kiriakidis, T. Pedersen, J. Kuhl, W. Koenig: Fabrication of photonic crystals by deep x-ray lithography, *Appl. Phys. Lett.* **71**(11), 1441–1443 (1997)
- 41.69 C. Cheng, A. Scherer: Fabrication of photonic band-gap crystals, *J. Vac. Sci. Technol. B* **13**(6), 2153–2153 (1995)
- 41.70 A. A. Zakhidov, R. H. Baughman, Z. Iqbal, C. Cui, I. Khayrullin, S. O. Dantas, J. Marti, V. G. Ralchenko: Carbon structures with three-dimensional periodicity at optical wavelengths, *Science* **282**, 897–901 (1998)
- 41.71 A. Blanco, E. Chomski, S. Grabtchak, M. Ibisate, S. John, S. W. Leonard, C. Lopez, F. Meseguer, H. Miguez, J. P. Mondia, G. A. Ozin, O. Toader, H. M. van Driel: Large-scale synthesis of a silicon photonic crystals with a complete three-dimensional bandgap near 1.5 micrometres, *Nature* **405**, 437–440 (2000)
- 41.72 J. E. G. J. Wijnhoven, W. L. Vos: Preparation of photonic crystals made of air spheres in titania, *Science* **281**, 802–804 (1998)
- 41.73 Y. Xia, B. Gates, Y. Yin, Y. Lu: Mono-dispersed colloidal spheres: Old materials with new applications, *Adv. Mater.* **12**(10), 693–713 (2000)
- 41.74 A. Stein: Sphere templating methods for periodic porous solids, *Micropor. Mesopor. Mater.* **44–45**, 227–239 (2001)
- 41.75 J. Martorell, N. M. Lawandy: Observation of inhibited spontaneous emission in a periodic dielectric structure, *Phys. Rev. Lett.* **65**(15), 1877–1880 (1990)
- 41.76 I. I. Tarhan, G. H. Watson: Photonic band structure of FCC colloidal crystals, *Phys. Rev. Lett.* **76**(2–8), 315–318 (1996)
- 41.77 Y. A. Vlasov, M. Deutsch, D. J. Norris: Single-domain spectroscopy of self-assembled photonic crystals, *Appl. Phys. Lett.* **76**(12), 1627–1629 (2000)
- 41.78 D. C. Reynolds, F. Lopez-Tejiera, D. Cassagne, F. Garcia-Vidal, C. Jouanin, J. Sanchez-Dehesa: Spectral properties of opal-based photonic crystals having a SiO₂ matrix, *Phys. Rev. B* **60**(16), 11422–11426 (1999)
- 41.79 V. N. Bogomolov, S. V. Gaponenko, I. N. Germanenko, A. M. Kapitonov, E. P. Petrov, N. V. Gaponenko, A. V. Prokofiev, A. N. Ponyavina, N. I. Silvanovich, S. M. Samoilovich: Photonic band gap phenomenon and optical properties of artificial opals, *Phys. Rev. E* **55**(6), 7619–7625 (1997)
- 41.80 S. G. Romanov, A. V. Fokin, R. M. De La Rue: Stop-band structure in complementary three-dimensional opal-based photonic crystals, *J. Phys. Condens. Matter* **11**, 3593–3600 (1999)
- 41.81 R. Biswas, M. M. Sigalas, G. Subramania, K. M. Ho: Photonic band gaps in colloidal systems, *Phys. Rev. B* **57**(9), 3701–3705 (1998)
- 41.82 A. Richel, N. P. Johnson, D. W. McComb: Observation of Bragg reflection in photonic crystals synthesized from air spheres in a titania matrix, *Appl. Phys. Lett.* **76**(14), 1816–1818 (2000)
- 41.83 B. T. Holland, C. F. Blanford, A. Stein: Synthesis of macroporous minerals with highly ordered three-dimensional arrays of spheroidal voids, *Science* **281**, 538–540 (1998)
- 41.84 M. S. Thijssen, R. Sprik, J. E. G. J. Wijnhoven, M. Megens, T. Narayanan, A. Lagendijk, W. L. Vos: Inhibited light propagation and broadband reflection in photonic air-sphere crystals, *Phys. Rev. Lett.* **83**(14), 2730–2733 (1999)
- 41.85 F. Meseguer, A. Blanco, H. Miguez, F. Garcia-Santamaria, M. Ibisate, C. Lopez: Synthesis of inverse opals, *Coll. Surf. A* **202**, 281–290 (2002)
- 41.86 O. D. Velev, E. W. Kaler: Research news: Structured porous materials via colloidal crystal templating: From inorganic oxides to metals, *Adv. Mater.* **12**(7), 531–534 (2000)
- 41.87 O. D. Velev, A. M. Lenhoff: Colloidal crystals as templates for porous materials, *Current Opin. Coll. Interf. Sci.* **5**, 56–63 (2000)
- 41.88 F. Blanford, H. Yan, R. C. Schroden, M. Al-Daous, A. Stein: Gems of chemistry and physics: Macroporous metal oxides with 3D order, *Adv. Mater.* **13**(6), 401–407 (2001)
- 41.89 A. M. Kapitonov, N. V. Gaponenko, V. N. Bogomolov, A. V. Prokofiev, S. M. Samoilovich, S. V. Gaponenko: Photonic stop band in a three-dimensional SiO₂/TiO₂ lattice, *Phys. Stat. Sol. (a)* **165**(1), 119–123 (1998)
- 41.90 N. Matsuura, S. Yang, P. Sun, H. E. Ruda: Development of highly-ordered, ferroelectric inverse opal films using sol-gel infiltration, *Appl. Phys. A* **81**, 379–384 (2005)
- 41.91 S. M. Yang, H. Miguez, G. A. Ozin: Opal circuits of light – planarized micro photonic crystals chips, *Adv. Funct. Mater.* **12**(6–7), 4254–4261 (2002)
- 41.92 A. Polman, P. Wiltzius: Materials science aspects of PCs, *MRS Bull.*, 608–610 (2001)
- 41.93 V. L. Colvin: From opals to optics: Colloidal photonic crystals, *MRS Bull.*, 637–641 (2001)
- 41.94 S. H. Park, D. Qin, Y. Xia: Crystallization of mesoscale particles over large areas, *Adv. Mater.* **10**(3), 1028–1032 (1998)
- 41.95 P. Jiang, J. Bertone, K. Hwang, V. Colvin: Single-crystal colloidal multi-layers of controlled thickness, *Chem. Mat.* **11**, 2132–2140 (1999)
- 41.96 Y. A. Vlasov, X.-Z. Bo, J. C. Sturm, D. J. Norris: On-chip natural assembly of silicon photonic bandgap crystals, *Nature* **414**, 289–293 (2001)

- 41.97 B. Gates, Y. Xia: Photonic crystals that can be addressed with an external magnetic field, *Adv. Mater.* **13**(21), 1605–1608 (2001)
- 41.98 I. Soten, H. Miguez, S. M. Yang, S. Petrov, N. Coombs, N. Tetreault, N. Matsuura, H. E. Ruda, G. A. Ozin: Barium titanate inverted opals – synthesis, characterization, and optical properties, *Adv. Funct. Mater.* **12**(1), 71–77 (2002)
- 41.99 M. C. Wanke, O. Lehmann, K. Muller, Q. Wen, M. Stuke: Laser rapid prototyping of photonic band-gap microstructures, *Science* **275**, 1284–1286 (1997)
- 41.100 B. H. Cumpston, S. P. Ananthavel, S. Barlow, D. L. Dyer, J. E. Ehrlich, L. L. Erskine, A. A. Heikal, S. M. Kuebler, I.-Y. S. Lee, D. McCord-Maughon, J. Qin, H. Rockel, M. Rumi, X.-L. Wu, S. R. Marder, J. W. Perry: Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication, *Nature* **398**, 51–54 (1999)
- 41.101 H.-B. Sun, S. Matsuo, H. Misawa: Three-dimensional photonic crystals structures achieved with two-photon-absorption photo-polymerization of resin, *Appl. Phys. Lett.* **74**(6), 786–788 (1999)
- 41.102 M. Campbell, D. Sharp, M. Harrison, R. Denning, A. Turberfield: Fabrication of photonic crystals for the visible spectrum by holographic lithography, *Nature* **404**, 53–56 (2000)
- 41.103 J. Martorell, R. Vilaseca, R. Corbalan: Second harmonic generation in a photonic crystal, *Appl. Phys. Lett.* **70**(6), 702–704 (1997)
- 41.104 H. Inouye, Y. Kanemitsu: Direct observation of non-linear effects in a one dimensional photonic crystal, *Appl. Phys. Lett.* **82**(8), 1155–1157 (2003)
- 41.105 Z.-Z. Gu, T. Iyoda, A. Fujishima, O. Sato: Photo reversible regulation of optical stop bands, *Adv. Mater.* **13**(7), 1295–1298 (2001)
- 41.106 X. Hu, Q. Zhang, Y. Liu, B. Cheng, D. Zhang: Ultrafast three-dimensional tunable photonic crystal, *Appl. Phys. Lett.* **83**(13), 2518–2520 (2003)
- 41.107 B. Li, J. Zou, X. J. Wang, X. H. Liu, J. Zi: Ferroelectric inverse opals with electrically tunable photonic band gap, *Appl. Phys. Lett.* **83**(23), 4704–4706 (2003)
- 41.108 T. B. Xu, Z. Y. Cheng, Q. M. Zhang, R. H. Baughman, C. Cui, A. A. Zakhidov, J. Su: Fabrication and characterization of three dimensional periodic ferroelectric polymer-silica opal composites and inverse opals, *J. Appl. Phys.* **88**(1), 405–409 (2000)
- 41.109 S. John, K. Busch: Photonic bandgap formation and tunability in certain self-organizing systems, *J. Lightwave Technol.* **17**(11), 1931–1943 (1999)
- 41.110 J. D. Joannopoulos: The almost-magical world of photonic crystals, *Braz. J. Phys.* **26**(1), 53–67 (1996)
- 41.111 X. Yoshino, Y. Kawagishi, M. Ozaki, A. Kose: Mechanical tuning of the optical properties of plastic opal as a photonic crystals, *Jpn. J. Appl. Phys. Pt. 2* **38**(7A), L786–78 (1999)
- 41.112 S. W. Leonard, H. M. van Driel, J. Schilling, R. B. Wehrspohn: Ultrafast band edge tuning of a two dimensional silicon photonic crystal via free carrier injection, *Phys. Rev. B* **66**, 161102–1–111102–4 (2002)
- 41.113 B. Gates, S. H. Park, Y. Xia: Tuning the photonic bandgap properties of crystalline arrays of polystyrene beads by annealing at elevated temperatures, *Adv. Mater.* **12**(9), 653–656 (2000)
- 41.114 P. A. Bermel, M. Warner: Photonic bandgap structure of highly deformable self-assembling systems, *Phys. Rev. E* **65**(1), 010702(R)–1–010702(R)–4 (2001)
- 41.115 Y. Iwayama, J. Yamanaka, Y. Takiguchi, M. Takasaka, K. Ito, T. Shinohara, T. Sawada, M. Yonese: Optically tunable gelled photonic crystal covering almost the entire visible light wavelength region, *Langmuir* **19**(4), 977–980 (2003)
- 41.116 S. Jun, Y.-S. Cho: Deformation-induced bandgap tuning of 2D silicon-based photonic crystals, *Opt. Express* **11**(21), 2769–2774 (2003)
- 41.117 C.-S. Kee, K. Kim, H. Lim: Tuning of anisotropic optical properties of 2D dielectric photonic crystals, *Physica B* **338**, 153–158 (2003)
- 41.118 K. Sumioka, H. Kayashima, T. Tsutsui: Tuning the optical properties of inverse opal photonic crystals by deformation, *Adv. Mater.* **14**(18), 1284–1286 (2002)
- 41.119 H. Miguez, F. Meseguer, C. Lopez, A. Blanco, J. S. Moya, J. Requena, A. Mifsud, V. Fornes: Control of the photonic crystal properties of fcc-packed sub-micrometer SiO₂ spheres by sintering, *Adv. Mater.* **10**(6), 480–483 (1998)
- 41.120 H. Pier, E. Kapon, M. Moser: Strain effects and phase transitions in photonic crystal resonator crystals, *Nature* **407**, 880–883 (2000)
- 41.121 N. Malkova, V. Gopalan: Strain tunable optical valves at T-junction waveguides in photonic crystals, *Phys. Rev. B* **68**, 245115–1–245115–6 (2003)
- 41.122 S. Kim, V. Gopalan: Strain tunable photonic band gap crystals, *Appl. Phys. Lett.* **78**(20), 3015–3017 (2001)
- 41.123 C. W. Wong, P. T. Rakich, S. G. Johnson, M. Qi, H. I. Smith, E. P. Ippen, L. C. Kimmerling, Y. Jeon, G. Barbastathis, S.-G. Kim: Strain-tunable silicon photonic bandgap microcavities in optical waveguides, *Appl. Phys. Lett.* **84**(8), 1242–1244 (2004)
- 41.124 S. H. Foulger, P. Jiang, A. Lattam, D. W. Smith, J. Ballato, D. E. Dausch, S. Grego, B. R. Stoner: Photonic crystal composites with reversible high-frequency stop band shifts, *Adv. Mater.* **15**(9), 685–689 (2003)
- 41.125 S. Rajic, J. L. Corbeil, P. G. Datskos: Feasibility of tunable MEMS photonic crystal devices, *Ultramicroscopy* **97**, 473–479 (2003)

- 41.126 N. Matsuura, H.E. Ruda, B.G. Yacobi: Configurable photonic device, US Patent 09/918398 [pending]
- 41.127 H. Hiller, J. Daleiden, C. Prott, F. Römer, S. Irmer, V. Rangelov, A. Tarraf, S. Schüler, M. Strassner: Potential for a micromachined actuation of ultra-wide continuously tunable optoelectronic devices, *Appl. Phys. B* **75**, 3–13 (2002)
- 41.128 W. Suh, S. Fan: Mechanically switchable photonic crystal filter with either all-pass transmission or flat-top reflection characteristics, *Opt. Lett.* **28**(19), 1763–1765 (2003)

42. Quantum Wells, Superlattices, and Band-Gap Engineering

This chapter reviews the principles of band-gap engineering and quantum confinement in semiconductors, with a particular emphasis on their optoelectronic properties. The chapter begins with a review of the fundamental principles of band-gap engineering and quantum confinement. It then describes the optical and electronic properties of semiconductor quantum wells and superlattices at a tutorial level, before describing the principal optoelectronic devices. The topics covered include edge-emitting lasers and light-emitting diodes (LEDs), resonant cavity LEDs and vertical-cavity surface-emitting lasers (VCSELs), quantum cascade lasers, quantum-well solar cells, superlattice avalanche photodiodes, inter-sub-band detectors, and quantum-well light modulators. The chapter concludes with a brief review of current research topics, including a discussion of quantum-dot structures.

42.1 Principles of Band-Gap Engineering and Quantum Confinement	1022
42.1.1 Lattice Matching	1022
42.1.2 Quantum-Confined Structures	1023

The need for efficient light-emitting diodes and lasers operating over the whole of the visible spectrum and also the fibre-optic windows at 1.3 μm and 1.55 μm drives research into new direct-gap semiconductors to act as the active materials. Since the emission wavelength of a semiconductor corresponds to its band-gap energy, research focuses on engineering new materials which have their band gaps at custom-designed energies. This science is called *band-gap engineering*.

In the early years of semiconductor optoelectronics, the band gaps that could be achieved were largely determined by the physical properties of key III–V materials such as GaAs and its alloys such as AlGaAs and InGaAs. Then in 1970 a major breakthrough occurred when *Esaki* and *Tsu* invented the semiconductor quantum well and superlattice [42.1]. They realised that developments in epitaxial crystal growth could open the door to new

42.2 Optoelectronic Properties of Quantum-Confined Structures	1024
42.2.1 Electronic States in Quantum Wells and Superlattices	1024
42.2.2 Interband Optical Transitions	1026
42.2.3 The Quantum-Confined Stark Effect	1028
42.2.4 Inter-Sub-Band Transitions	1029
42.2.5 Vertical Transport	1030
42.2.6 Carrier Capture and Relaxation	1031
42.3 Emitters	1032
42.3.1 Interband Light-Emitting Diodes and Lasers	1032
42.3.2 Quantum Cascade Lasers	1033
42.4 Detectors	1034
42.4.1 Solar Cells	1034
42.4.2 Avalanche Photodiodes	1034
42.4.3 Inter-Sub-Band Detectors	1035
42.4.4 Unipolar Avalanche Photodiodes	1035
42.5 Modulators	1036
42.6 Future Directions	1037
42.7 Conclusions	1038
References	1038

structures that exploit the principles of quantum confinement to engineer electronic states with custom-designed properties. They foresaw that these quantum-confined structures would be of interest both to research scientists, who would be able to explore uncharted areas of fundamental physics, and also to engineers, who would learn to use their unique properties for device applications. Their insight paved the way for a whole new breed of devices that are now routinely found in a host of everyday applications ranging from compact-disc players to traffic lights.

The emphasis of the chapter is on the optoelectronic properties of quantum-well and superlattice structures. We begin by outlining the basic principles of band-gap engineering and quantum confinement. We will then discuss the electronic states in quantum-confined structures and the optical properties that follow from

them. In Sects. 42.3–42.5 we will explain the principles of the main optoelectronic devices that employ quantum wells and superlattices, namely emitters, detectors and modulators. Finally we will indicate a few interesting recent developments that offer exciting prospects for future devices before drawing the chapter to its

conclusion. A number of texts cover these topics in more detail (e.g. [42.2–5]), and the interested reader is referred to these sources for a more comprehensive treatment. A description of the purely electronic properties of low-dimensional structures may be found in [42.6].

42.1 Principles of Band-Gap Engineering and Quantum Confinement

42.1.1 Lattice Matching

The art of band-gap engineering relies heavily on developments in the science of crystal growth. Bulk crystals grown from the melt usually contain a large number of impurities and defects, and optoelectronic devices are therefore grown by epitaxial methods such as liquid-phase epitaxy (LPE), molecular-beam epitaxy (MBE) and metalorganic vapour-phase epitaxy (MOVPE), which is also called metalorganic chemical vapour deposition (MOCVD) (Chapt. 14). The basic principle of epitaxy is to grow thin layers of very high purity on top of a bulk crystal called the substrate. The system is said to be *lattice-matched* when the lattice constants of the epitaxial layer and the substrate are identical. Lattice-matching reduces the number of dislocations in the epitaxial layer, but it also imposes tight restrictions on the band gaps that can be engineered easily, because there are only a relatively small number of convenient substrate materials available.

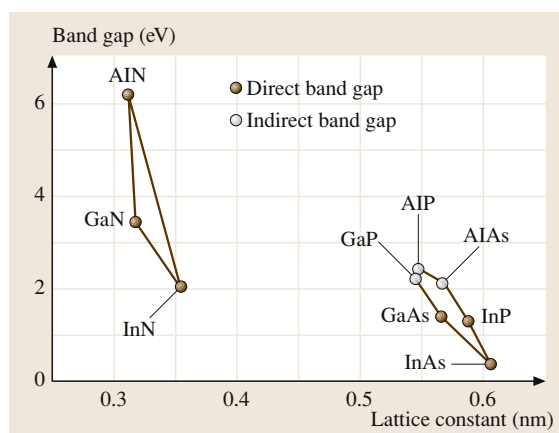


Fig. 42.1 Room-temperature band gap of a number of important optoelectronic III–V materials versus their lattice constant

We can understand this point more clearly by reference to Fig. 42.1. This diagram plots the band-gap energy E_g of a number of important III–V semiconductors as a function of their lattice constant. The majority of optoelectronic devices for the red/near-infrared spectral regions are either grown on GaAs or InP substrates. The simplest case to consider is an epitaxial layer of GaAs grown on a GaAs substrate, which gives an emission wavelength of 873 nm (1.42 eV). This wavelength is perfectly acceptable for applications involving short-range transmission down optical fibres. However, for long distances we require emission at 1.3 μm or 1.55 μm , while for many other applications we require emission in the visible spectral region.

Let us first consider the preferred fibre-optic wavelengths of 1.3 μm and 1.55 μm . There are no binary semiconductors with band gaps at these wavelengths, and so we have to use alloys to tune the band gap by varying the composition (Chapt. 31). A typical example is the ternary alloy $\text{Ga}_x\text{In}_{1-x}\text{As}$, which is lattice-matched to InP when $x = 47\%$, giving a band gap of 0.75 eV (1.65 μm). $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ photodiodes grown on InP substrates make excellent detectors for 1.55- μm radiation, but to make an emitter at this wavelength, we have to increase the band gap while maintaining the lattice-matching condition. This is achieved by incorporating a fourth element into the alloy – typically Al or P, which gives an extra design parameter that permits band-gap tuning while maintaining lattice-matching. Thus the quaternary alloys $\text{Ga}_{0.27}\text{In}_{0.73}\text{As}_{0.58}\text{P}_{0.42}$ and $\text{Ga}_{0.40}\text{In}_{0.60}\text{As}_{0.85}\text{P}_{0.18}$ give emission at 1.3 μm and 1.55 μm , respectively, and are both lattice-matched to InP substrates.

Turning now to the visible spectral region, it is a convenient coincidence that the lattice constants of GaAs and AlAs are almost identical. This means that we can grow relatively thick layers of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ on GaAs substrates without introducing dislocations and other defects. The band gap of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ varies quadratically

with x according to:

$$E_g(x) = (1.42 + 1.087x + 0.438x^2) \text{ eV}, \quad (42.1)$$

but unfortunately the gap becomes indirect for $x > 43\%$. We can therefore engineer direct band gaps of 1.42–1.97 eV, giving emission from 873 nm in the near infrared to 630 nm in the red spectral range. Much work has been done on quaternary alloys such as AlGaInP, (Chapt. 31) but it has not been possible to make blue- and green-emitting devices based on GaAs substrates to date, due to the tendency for arsenic and phosphorous compounds to become indirect as the band gap increases.

The approach for the blue end of the spectrum preferred at present is to use nitride-based compounds. (Chapt. 32) Early work on nitrides established that their large direct gaps made them highly promising candidates for use as blue/green emitters [42.7]. However, it was not until the 1990s that this potential was fully realised. The rapid progress followed two key developments, namely the activation of p-type dopants and the successful growth of strained $\text{In}_x\text{Ga}_{1-x}\text{N}$ quantum wells which did not satisfy the lattice-matching condition [42.8]. The second point goes against the conventional wisdom of band-gap engineering and highlights the extra degrees of freedom afforded by quantum-confined structures, as will now be discussed.

42.1.2 Quantum-Confined Structures

A quantum-confined structure is one in which the motion of the electrons (and/or holes) are confined in one or more directions by potential barriers. The general scheme for classifying quantum-confined structures is given in Table 42.1. In this chapter we will be concerned primarily with *quantum wells*, although we will briefly refer to *quantum wires* and *quantum dots* in Sect. 42.6. Quantum size effects become important when the thickness of the layer becomes comparable with the de Broglie wavelength of the electrons or holes.

Table 42.1 Classification of quantum-confined structures. In the case of quantum wells, the confinement direction is usually taken as the z -axis

Structure	Confined directions	Free directions (dimensionality)
Quantum well	1 (z)	2 (x, y)
Quantum wire	2	1
Quantum dot (or box)	3	none

If we consider the free thermal motion of a particle of mass m in the z -direction, the de Broglie wavelength at a temperature T is given by

$$\lambda_{\text{deB}} = \frac{h}{\sqrt{mk_B T}}. \quad (42.2)$$

For an electron in GaAs with an effective mass of $0.067m_0$, we find $\lambda_{\text{deB}} = 42 \text{ nm}$ at 300 K. This implies that we need structures of thickness $\approx 10 \text{ nm}$ in order to observe quantum-confinement effects at room temperature. Layers of this thickness are routinely grown by the MBE or MOVPE techniques described in Chapt. 14.

Figure 42.2 shows a schematic diagram of a GaAs/AlGaAs quantum well. The quantum confinement is provided by the discontinuity in the band gap at the interfaces, which leads to a spatial variation of the conduction and valence bands, as shown in the lower half of the figure. The Al concentration is typically chosen to be around 30%, which gives a band-gap discontinuity of 0.36 eV according to (42.1). This splits roughly 2 : 1 between the conduction and valence bands, so that electrons see a confining barrier of 0.24 eV and the holes see 0.12 eV.

If the GaAs layers are thin enough, according to the criterion given above, the motion of the electrons and holes will be quantised in the growth (z) direction, giving

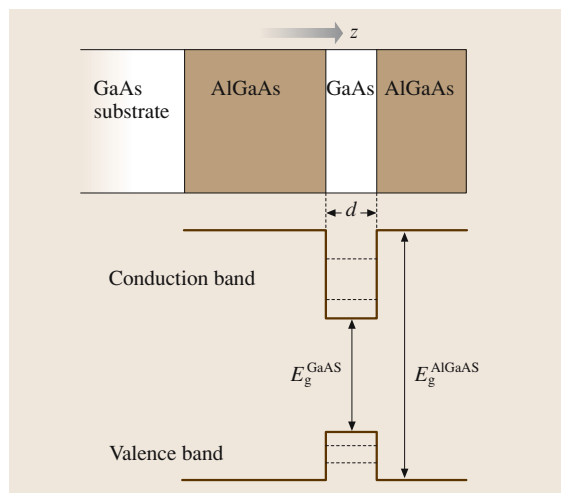


Fig. 42.2 Schematic diagram of the growth layers and resulting band diagram for a GaAs/AlGaAs quantum well of thickness d . The quantised levels in the quantum well are indicated by the *dashed lines*. Note that in real structures a GaAs buffer layer is usually grown immediately above the GaAs substrate

rise to a series of discrete energy levels, as indicated by the dashed lines inside the quantum well in Fig. 42.2. The motion is in the other two directions (i.e. the x - y plane) is still free, and so we have quasi two-dimensional (2-D) behaviour.

The quantisation of the motion in the z -direction has three main consequences. Firstly, the quantisation energy shifts the effective band edge to higher energy, which provides an extra degree of freedom in the art of band-gap engineering. Secondly, the confinement keeps the electrons and holes closer together and hence increases the radiative recombination probability. Finally, the density of states becomes independent of energy, in contrast to three-dimensional (3-D) materials, where the density of states is proportional to $E^{1/2}$. Many of the useful properties of the quantum wells follow from these three properties.

The arrangement of the bands shown in Fig. 42.2 in which both the electrons and holes are confined in the quantum well is called *type I band alignment*. Other types of band alignments are possible in which only one of the carrier types are confined (*type II band alignment*). Furthermore, the flexibility of the MBE and MOVPE growth techniques easily allows the growth of *superlattice* (SL) structures containing many repeated quantum wells with thin barriers separating them, as shown in Fig. 42.3. Superlattices behave like artificial one-dimensional periodic crystals, in which the periodicity is designed into the structure by the repetition of the quantum wells. The electronic states of SLs form delocalised *minibands* as the wave functions in neighbouring wells couple together through the thin barrier that separates them. Structures containing a smaller number of repeated wells or with thick barriers that prevent coupling between adjacent wells are simply called *multiple quantum well* (MQW) structures.

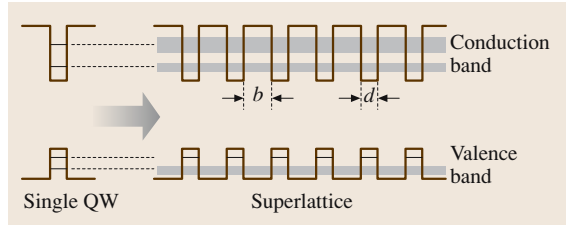


Fig. 42.3 Schematic diagram of a superlattice, showing the formation of minibands from the energy levels of the corresponding single quantum well (QW). The structure forms an artificial one-dimensional crystal with period $(d+b)$, where d and b represent the thickness of the QW and barrier regions respectively. The width of the minibands depends on the strength of the coupling through the barriers. It is frequently the case that the lowest hole states do not couple strongly, and hence remain localised within their respective wells, as shown in the figure

In the next section we will describe in more detail the electronic properties of quantum wells and superlattices. Before doing so, it is worth highlighting two practical considerations that are important additional factors that contribute to their usefulness. The first is that band-gap tunability can be achieved without using alloys as the active material, which is desirable because alloys inevitably contain more defects than simple compounds such as GaAs. The second factor is that the quantum wells can be grown as strained layers on top of a lattice with a different cell constant. A typical example is the $\text{In}_x\text{Ga}_{1-x}\text{N}/\text{GaN}$ quantum wells mentioned above. These layers do not satisfy the lattice-matching condition, but as long as the total $\text{In}_x\text{Ga}_{1-x}\text{N}$ thickness is less than the critical value, there is an energy barrier to the formation of dislocations. In practice this allows considerable extra flexibility in the band-gap engineering that can be achieved.

42.2 Optoelectronic Properties of Quantum-Confined Structures

42.2.1 Electronic States in Quantum Wells and Superlattices

Quantum Wells

The electronic states of quantum wells can be understood by solving the Schrödinger equation for the electrons and holes in the potential wells created by the band discontinuities. The simplest approach is the infinite-well model shown in Fig. 42.4a. The Schrödinger equation in

the well is

$$-\frac{\hbar^2}{2m_w^*} \frac{d^2\psi(z)}{dz^2} = E\psi(z), \quad (42.3)$$

where m_w^* is the effective mass in the well and z is the growth direction. Since the potential barriers are infinite, there can be no penetration into the barriers, and we must therefore have $\psi(z) = 0$ at the interfaces. If we choose our origin such that the quantum well runs from $z = 0$

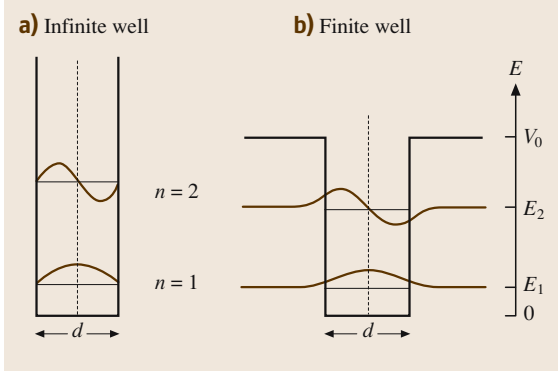


Fig. 42.4a,b Confined states in a quantum well of width d . **(a)** A perfect quantum well with infinite barriers. **(b)** A finite well with barriers of height V_0 . The wave functions for the $n = 1$ and $n = 2$ levels are sketched for both types of well

to $z = d$, d being the width of the well, the normalised wave functions take the form (see e.g. [42.9]):

$$\psi_n(z) = \sqrt{2/d} \sin k_n z, \quad (42.4)$$

where $k_n = (n\pi/d)$ and the quantum number n is an integer (≥ 1). The energy E_n is given by

$$E_n = \frac{\hbar^2 k_n^2}{2m_w^*} = \frac{\hbar^2}{2m_w^*} \left(\frac{n\pi}{d} \right)^2. \quad (42.5)$$

The wave functions of the $n = 1$ and $n = 2$ levels are sketched in Fig. 42.4a.

Although the infinite-well model is very simplified, it nonetheless provides a good starting point for understanding the general effects of quantum confinement. Equation (42.5) shows us that the energy is inversely proportional to d^2 , implying that narrow wells have larger confinement energies. Furthermore, the confinement energy is inversely proportional to the effective mass, which means that lighter particles experience larger effects. This also means that the heavy- and light-hole states have different energies, in contrast to bulk semiconductors in which the two types of hole states are degenerate at the top of the valence band.

Now let us consider the more realistic finite-well model shown in Fig. 42.4b. The Schrödinger equation in the well is unchanged, but in the barrier regions we now have:

$$-\frac{\hbar^2}{2m_b^*} \frac{d^2 \psi(z)}{dz^2} + V_0 \psi(z) = E \psi(z), \quad (42.6)$$

where V_0 is the potential barrier and m_b^* is the effective mass in the barrier. The boundary conditions require that

the wave function and particle flux $(1/m^*)d\psi/dz$ must be continuous at the interface. This gives a series of even and odd parity solutions which satisfy

$$\tan(kd/2) = \frac{m_w^* \kappa}{m_b^* k}, \quad (42.7)$$

and

$$\tan(kd/2) = -\frac{m_b^* k}{m_w^* \kappa}, \quad (42.8)$$

respectively. k is the wave vector in the well, given by

$$\frac{\hbar^2 k^2}{2m_w^*} = E_n, \quad (42.9)$$

while κ is the exponential decay constant in the barrier, given by

$$\frac{\hbar^2 \kappa^2}{2m_b^*} = V_0 - E_n. \quad (42.10)$$

Solutions to (42.7) and (42.8) are easily found by simple numerical techniques [42.9]. As with the infinite well, the eigenstates are labelled by the quantum number n and have parities of $(-1)^{n+1}$ with respect to the axis of symmetry about the centre of the well. The wave functions are approximately sinusoidal inside the well, but decay exponentially in the barriers, as illustrated in Fig. 42.4b. The eigen-energies are smaller than those of the infinite well due to the penetration of the barriers, which means that the wave functions are less well confined. There is only a limited number of solutions, but there is always at least one, no matter how small V_0 might be.

As an example we consider a typical GaAs/Al_{0.3}Ga_{0.7}As quantum well with $d = 10$ nm. The confinement energy is 245 meV for the electrons and 125 meV for the holes. The infinite well model predicts $E_1 = 56$ meV and $E_2 = 224$ meV for the electrons, whereas (42.7), (42.8) give $E_1 = 30$ meV and $E_2 = 113$ meV. For the heavy (light) holes the infinite-well model predicts 11 meV (40 meV) and 44 meV (160 meV) for the first two bound states, instead of the more accurate values of 7 meV (21 meV) and 29 meV (78 meV) calculated from the finite-well model. Note that the separation of the electron levels is greater than $k_B T$ at 300 K, so that the quantisation effects will be readily observable at room temperature.

Strained Quantum Wells

Even more degrees of freedom for tailoring the electronic states can be achieved by epitaxially stacking semiconductor layers with different lattice constants to form *strained quantum wells*. Examples include

$\text{In}_x\text{Ga}_{1-x}\text{As}$ on GaAs, and $\text{Si}_{1-x}\text{Ge}_x$ on Si. Large biaxial strain develops within the x - y plane of a quantum well grown on a substrate with a different lattice constant. In order to avoid the buildup of misfit dislocations at the interfaces, the strained layers need to be thinner than a certain critical dimension. For example, a defect-free strained $\text{In}_x\text{Ga}_{1-x}\text{As}$ layer on GaAs requires a thickness less than around 10 nm when $x = 0.2$. Since the band gap is related to the lattice constant, the strain induces a shift of the band edges which, in turn, affects many other properties. It is due to some of these effects that strained QW structures have become widely exploited in optoelectronic devices. (Chapt. 37)

The most significant effect of the strain is to alter the band gap and remove the valence-band degeneracy near the Γ valley. The splitting of the valence band is a consequence of the lattice distortion, which reduces the crystal symmetry from cubic to tetragonal [42.10]. There are essentially two types of strain. Compressive strain occurs when the active layer has a larger lattice constant than the substrate, for example in $\text{In}_x\text{Ga}_{1-x}\text{As}$ on GaAs. In this case, the band gap increases and the effective mass of the highest hole band decreases, while that of next valence band increases. The opposite case is that of tensile strain, which occurs when the active layer has a smaller lattice constant than the substrate, such as $\text{Si}_{1-x}\text{Ge}_x$ on Si. The ordering of the valence bands is opposite to the case of compressive strain, and the overall band gap is reduced.

Superlattices

The analytical derivation of the allowed energy values in a superlattice (SL) is similar to that for a single QW, with the appropriate change of the boundary conditions imposed by the SL periodicity. The mathematical description of a superlattice is similar to a one-dimensional crystal lattice, which allows us to borrow the formalism of the band theory of solids, including the well-known Kronig-Penney model [42.9]. Within this model, the electron envelope wave function $\psi(z)$ can be expressed as a superposition of Bloch waves propagating along the z -axis. For a SL with a barrier height V_0 , the allowed energy is calculated numerically as a solution of the transcendental equation involving the Bloch wave vector:

$$\begin{aligned} \cos(ka) &= \cos(kd) \cos(\kappa'b) \\ &\quad - \frac{k^2 + \kappa'^2}{2k\kappa'} \sin(kd) \sin(\kappa'b), \\ E &> V_0, \end{aligned} \quad (42.11)$$

$$\begin{aligned} \cos(ka) &= \cos(kd) \cos(\kappa b) \\ &\quad - \frac{k^2 - \kappa^2}{2k\kappa} \sin(kd) \sin(\kappa b), \\ E &< V_0, \end{aligned} \quad (42.12)$$

where $a \equiv (b + d)$ is the period, and k and κ are given by (42.9) and (42.10), respectively. The decay constant κ' is given by:

$$E - V_0 = \frac{\hbar^2 \kappa'^2}{2m_b^*}. \quad (42.13)$$

The electronic states in superlattices can be understood in a more qualitative way by reference to Fig. 42.3 and making use of the analogy with the tight-binding model of band formation in solids. Isolated atoms have discrete energy levels which are localised on the individual atom sites. When the atoms are brought close together, the energy levels broaden into bands, and the overlapping wave functions develop into extended states. In the same way, repeated quantum-well structures with large values of the barrier thickness b (i. e. MQWs) have discrete levels with wave functions localised within the wells. As the barrier thickness is reduced, the wave functions of adjacent wells begin to overlap and the discrete levels broaden into minibands. The wave functions in the minibands are delocalised throughout the whole superlattice. The width of the miniband depends on the cross-well coupling, which is determined by the barrier thickness and the decay constant κ (42.10). In general, the higher-lying states give rise to broader minibands because κ decreases with E_n . Also, the heavy-hole minibands are narrower than the electron minibands, because the cross-well coupling decreases with increasing effective mass.

42.2.2 Interband Optical Transitions

Absorption

The optical transitions in quantum wells take place between electronic states that are confined in the z -direction but free in the x - y plane. The transition rate can be calculated from Fermi's golden rule, which states that the probability for optical transitions from the initial state $|i\rangle$ at energy E_i to the final state $|f\rangle$ at energy E_f is given by:

$$W(i \rightarrow f) = \frac{2\pi}{\hbar} |\langle f | \mathbf{e} \cdot \mathbf{E} | i \rangle|^2 g(\hbar\omega), \quad (42.14)$$

where \mathbf{e} is the electric dipole of the electron, \mathbf{E} is the electric field of the light wave, and $g(\hbar\omega)$ is the joint density of states at photon energy $\hbar\omega$. Conservation of

energy requires that $E_f = (E_i + \hbar\omega)$ for absorption, and $E_f = (E_i - \hbar\omega)$ for emission.

Let us consider a transition from a confined hole state in the valence band with quantum number n to a confined electron state in the conduction band with quantum number n' . We apply Bloch's theorem to write the wave functions in the following form:

$$\begin{aligned} |i\rangle &= u_v(\mathbf{r}) \exp(i\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}) \psi_{hn}(z) \\ |f\rangle &= u_c(\mathbf{r}) \exp(i\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}) \psi_{en'}(z), \end{aligned} \quad (42.15)$$

where $u_v(\mathbf{r})$ and $u_c(\mathbf{r})$ are the envelope function for the valence and conduction bands, respectively, \mathbf{k}_{xy} is the in-plane wave vector for the free motion in the x - y plane, \mathbf{r}_{xy} being the xy component of the position vector, and $\psi_{hn}(z)$ and $\psi_{en'}(z)$ are the wave functions for the confined hole and electron states in the z -direction. We have applied conservation of momentum here so that the in-plane wave vectors of the electron and hole are the same.

On inserting these wave functions into (42.14) we find that the transition rate is proportional to both the square of the overlap of the wave functions and the joint density of states [42.11]:

$$W \propto |\langle \psi_{en'}(z) | \psi_{hn}(z) \rangle|^2 g(\hbar\omega). \quad (42.16)$$

The wave functions of infinite wells are orthogonal unless $n = n'$, which gives a selection rule of $\Delta n = 0$. For finite wells, the $\Delta n = 0$ selection rule is only approximately obeyed, although transitions between states of different parity (i. e. Δn odd) are strictly forbidden. The joint density of states is independent of energy due to the quasi-2-D nature of the quantum well.

Figure 42.5a illustrates the first two strong transitions in a typical quantum well. These are the $\Delta n = 0$ transitions between the first and second hole and electron levels. The threshold energy for these transitions is equal to

$$\hbar\omega = E_g + E_{hn} + E_{en}. \quad (42.17)$$

The lowest value is thus equal to $(E_g + E_{h1} + E_{e1})$, which shows that the optical band gap is shifted by the sum of the electron and hole confinement energies. Once the photon energy exceeds the threshold set by (42.17), a continuous band of absorption occurs with the absorption coefficient independent of energy due to the constant 2-D density of states of the quantum well. The difference between the absorption of an ideal quantum well with infinite barriers and the equivalent bulk semiconductor is illustrated in Fig. 42.5b. In the quantum well we find a series of steps with constant absorption coefficients, whereas in the bulk the absorption varies as

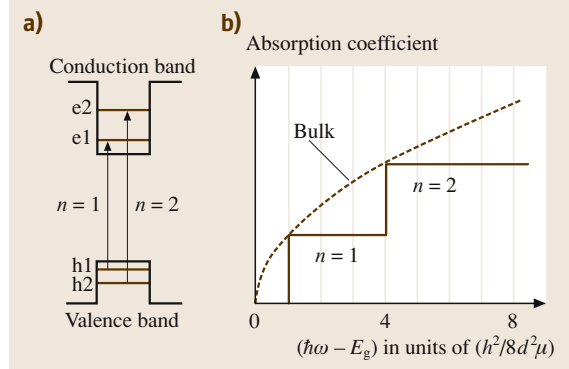


Fig. 42.5a,b Interband optical transitions between confined states in a quantum well. **(a)** Schematic diagram showing the $\Delta n = 0$ transitions between the $n = 1$ and $n = 2$ sub-bands. **(b)** Absorption spectrum for an infinite quantum well of thickness d with a reduced electron-hole mass μ in the absence of excitonic effects. The absorption spectrum of the equivalent bulk semiconductor is shown by the dashed line for comparison

$(\hbar\omega - E_g)^{1/2}$ for $\hbar\omega > E_g$. Thus the transition from 3-D to 2-D alters the shape of the absorption curve, and also causes an effective shift in the band gap by $(E_{h1} + E_{e1})$.

Up to this point, we have neglected the Coulomb interaction between the electrons and holes which are involved in the transition. This attraction leads to the formation of bound electron-hole pairs called excitons. The exciton states of a quantum well can be modelled as 2-D hydrogen atoms in a material with relative dielectric constant ϵ_r . In this case, the binding energy E^X is given by [42.12]:

$$E^X(v) = \frac{\mu}{m_0} \frac{1}{\epsilon_r^2} \frac{1}{(v - 1/2)^2} R_H, \quad (42.18)$$

where v is an integer ≥ 1 , m_0 is the electron mass, μ is the reduced mass of the electron-hole pair, and R_H is the Rydberg constant for hydrogen (13.6 eV). This contrasts with the standard formula for 3-D semiconductors in which E^X varies as $1/v^2$ rather than $1/(v - 1/2)^2$, and implies that the binding energy of the ground-state exciton is four times larger in 2-D than 3-D. This allows excitonic effects to be observed at room temperature in quantum wells, whereas they are only usually observed at low temperatures in bulk semiconductors.

Figure 42.6 compares the band-edge absorption of a GaAs MQW sample with that of bulk GaAs at room temperature [42.13]. The MQW sample contained 77 GaAs quantum wells of thickness 10 nm with thick $\text{Al}_{0.28}\text{Ga}_{0.72}\text{As}$ barriers separating them. The shift of

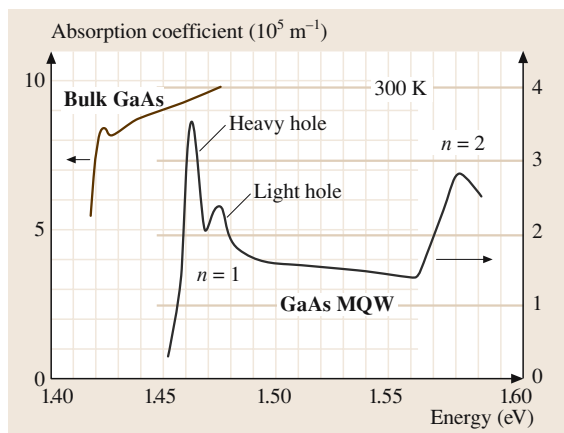


Fig. 42.6 Absorption spectrum of a 77-period GaAs/Al_{0.28}Ga_{0.72}As MQW structure with 10-nm quantum wells at room temperature. The absorption spectrum of bulk GaAs is included for comparison. (After [42.13], © 1982 AIP)

the band edge of the MQW to higher energy is clearly observed, together with the series of steps due to each $\Delta n = 0$ transition. The sharp lines are due to excitons, which occur at energies given by

$$\hbar\omega = E_g + E_{hh1} + E_{el} - E^X. \quad (42.19)$$

Equation (42.18) predicts that E^X should be around 17 meV for the ground-state exciton of an ideal GaAs quantum well, compared to 4.2 meV for the bulk. The actual QW exciton binding energies are somewhat smaller due to the tunnelling of the electrons and holes into the barriers, with typical values of around 10 meV. However, this is still substantially larger than the bulk value and explains why the exciton lines are so much better resolved for the QW than the bulk. The absorption spectrum of the QW above the exciton lines is approximately flat due to the constant density of states in 2-D, which contrasts with the rising absorption of the bulk due to the parabolic 3-D density of states. Separate transitions are observed for the heavy and light holes. This follows from their different effective masses, and can also be viewed as a consequence of the lower symmetry of the QW compared to the bulk.

Emission

Emissive transitions occur when electrons excited in the conduction band drop down to the valence band and recombine with holes. The optical intensity $I(\hbar\omega)$ is proportional to the transition rate given by (42.14) multiplied by the probability that the initial state is occupied

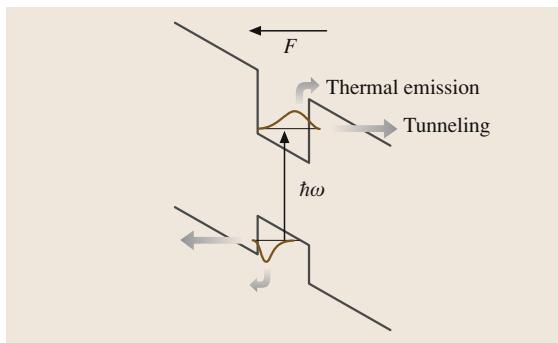


Fig. 42.7 Schematic band diagram and wave functions for a quantum well in a DC electric field F

and the final state is empty:

$$I(\hbar\omega) \propto W(c \rightarrow v) f_c (1 - f_v), \quad (42.20)$$

where f_c and f_v are the Fermi–Dirac distribution functions in the conduction and valence bands, respectively. In thermal equilibrium, the occupancy of the states is largest at the bottom of the bands and decays exponentially at higher energies. Hence the luminescence spectrum of a typical GaAs QW at room temperature usually consists of a peak of width $\sim k_B T$ at the effective band gap of $(E_g + E_{hh1} + E_{el})$. At lower temperatures the spectral width is affected by inhomogeneous broadening due to unavoidable fluctuations in the well thickness. Furthermore, in quantum wells employing alloy semiconductors, the microscopic fluctuations in the composition can lead to additional inhomogeneous broadening. This is particularly true of InGa_N/Ga_N quantum wells, where indium compositional fluctuations produce substantial inhomogeneous broadening even at room temperature.

The intensity of the luminescence peak in quantum wells is usually much larger than that of bulk materials because the electron–hole overlap is increased by the confinement. This leads to faster radiative recombination, which then wins out over competing nonradiative decay mechanisms and leads to stronger emission. This enhanced emission intensity is one of the main reasons why quantum wells are now so widely used in diode lasers and light-emitting diodes.

42.2.3 The Quantum-Confined Stark Effect

The *quantum-confined Stark effect* (QCSE) describes the response of the confined electron and hole states in quantum wells to a strong direct-current (DC) electric field applied in the growth (z) direction. The field is

usually applied by growing the quantum wells inside a p–n junction, and then applying reverse bias to the diode. The magnitude of the electric field F is given by:

$$F = \frac{V^{\text{built-in}} - V^{\text{bias}}}{L_i}, \quad (42.21)$$

where $V^{\text{built-in}}$ is the built-in voltage of the diode, V^{bias} is the bias voltage, and L_i is the total thickness of the intrinsic region. $V^{\text{built-in}}$ is approximately equal to the band-gap voltage of the doped regions (≈ 1.5 V for a GaAs diode).

Figure 42.7 gives a schematic band diagram of a quantum well with a strong DC electric field applied. The field tilts the potential and distorts the wave functions as the electrons tend to move towards the anode and the holes towards the cathode. This has two important consequences for the optical properties. Firstly, the lowest transition shifts to lower energies due to the electrostatic interaction between the electric dipole induced by the field and the field itself. At low fields the dipole is proportional to F , and the red shift is thus proportional to F^2 (the quadratic Stark effect). At higher fields, the dipole saturates at a value limited by ed , where e is the electron charge and d the well width, and the Stark shift is linear in F . Secondly, the parity selection rule no longer applies due to the breaking of the inversion symmetry about the centre of the well. This means that *forbidden* transitions with Δn equal to an odd number become allowed. At the same time, the $\Delta n = 0$ transitions grad-

ually weaken with increasing field as the distortion to the wave functions reduces the electron–hole overlap.

Figure 42.8 shows the normalised room-temperature photocurrent spectra of a GaAs/Al_{0.3}Ga_{0.7}As MQW p–i–n diode containing 9.0-nm quantum wells at 0 V and –10 V applied bias. These two bias values correspond to field strengths of around 15 kV/cm and 115 kV/cm respectively. The photocurrent spectrum closely resembles the absorption spectrum, due to the field-induced escape of the photoexcited carriers in the QWs into the external circuit (Sect. 42.2.5). The figure clearly shows the Stark shift of the absorption edge at the higher field strength, with a red shift of around 20 meV (≈ 12 nm) at –10 V bias for the hh1 \rightarrow e1 transition. The intensity of the line weakens somewhat due to the reduction in the electron–hole overlap, and there is lifetime broadening caused by the field-assisted tunnelling. Several parity-forbidden transitions are clearly observed. The two most obvious ones are identified with arrows, and correspond to the hh2 \rightarrow e1 and hh1 \rightarrow e2 transitions, respectively.

A striking feature in Fig. 42.8 is that the exciton lines are still resolved even at very high field strengths. In bulk GaAs the excitons ionise at around 5 kV/cm [42.11], but in QWs the barriers inhibit the field ionisation, and excitonic features can be preserved even up to ≈ 300 kV/cm [42.14]. The ability to control the absorption spectrum by the QCSE is the principle behind a number of important modulator devices, which will be discussed in Sect. 42.5.

In the case of a superlattice, such as that illustrated in Fig. 42.3, a strong perpendicular electric field can break the minibands into discrete energy levels local to each QW, due to the band-gap tilting effect represented in Fig. 42.7. The possibility of using an electric field to modify the minibands of a superlattice is yet another remarkable ability of band-gap engineering to achieve control over the electronic properties by directly using fundamental principles of quantum mechanics.

42.2.4 Inter-Sub-Band Transitions

The engineered band structure of quantum wells leads to the possibility of *inter-sub-band (ISB) transitions*, which take place between confined states within the conduction or valence bands, as illustrated schematically in Fig. 42.9. The transitions typically occur in the infrared spectral region. For example, the e1 \leftrightarrow e2 ISB transition in a 10-nm GaAs/AlGaAs quantum well occurs at around 15 μm . For ISB absorption transitions we must first dope the conduction band so that there is a large population of electrons in the e1 level, as shown

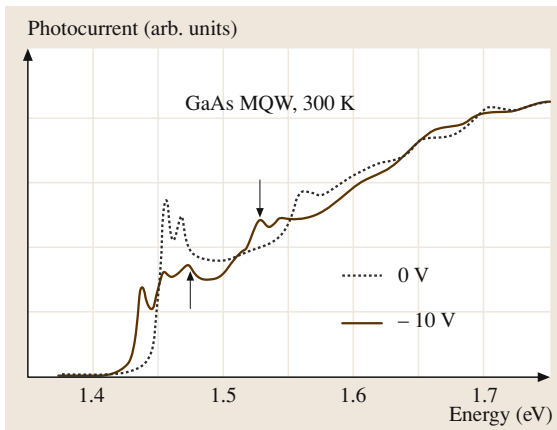


Fig. 42.8 Room-temperature photocurrent spectra for a GaAs/Al_{0.3}Ga_{0.7}As MQW p–i–n diode with a 1- μm -thick i-region at zero bias and –10 V. The quantum well thickness was 9.0 nm. The arrows identify transitions that are forbidden at zero field. (After [42.14], © 1991 IEEE)

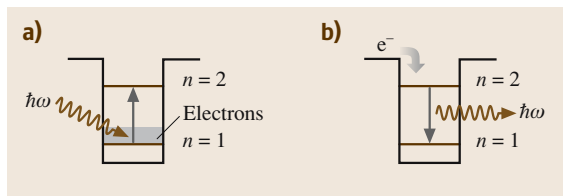


Fig. 42.9 (a) Inter-sub-band absorption in an n-type quantum well. (b) Inter-sub-band emission following electron injection to a confined upper level in the conduction band

in Fig. 42.9a. This is typically achieved by n-type doping of the barriers, which produces a large electron density as the extrinsic electrons drop from the barriers to the confined states in the QW. Undoped wells are used for ISB emission transitions, and electrons must first be injected into excited QW states, as shown in Fig. 42.9b.

The basic properties of ISB transitions can be understood by extension of the principles outlined in Sect. 42.2.2. The main difference is that the envelope functions for the initial and final states are the same, since both states lie in the same band. The transition rate for a conduction-band ISB transition then turns out to be given by:

$$W^{\text{ISB}} \propto |\langle \psi_{en'}(z) | z | \psi_{en}(z) \rangle|^2 g(\hbar\omega), \quad (42.22)$$

where n and n' are the quantum numbers for the initial and final confined levels. The z operator within the Dirac bracket arises from the electric-dipole interaction and indicates that the electric field of the light wave must be parallel to the growth direction. Furthermore, the odd parity of the z operator implies that the wave functions must have different parities, and hence that Δn must be an odd number. The use of ISB transitions in infrared emitters and detectors will be discussed in Sects. 42.3.2, 42.4.3 and 42.4.4.

42.2.5 Vertical Transport

Quantum Wells

Vertical transport refers to the processes by which electrons and holes move in the growth direction. Issues relating to vertical transport are important for the efficiency and frequency response of most QW optoelectronic devices. The transport is generally classified as either *bipolar*, when both electrons and holes are involved, or *unipolar*, when only one type of carrier (usually electrons) is involved. In this section we will concentrate primarily on bipolar transport in QW detectors and QCSE modulators. Bipolar transport in light-emitting devices is discussed in Sects. 42.2.6

and 42.3.1, while unipolar transport in quantum cascade lasers is discussed in Sect. 42.3.2.

In QW detectors and QCSE modulators the diodes are operated in reverse bias. This produces a strong DC electric field and tilts the bands as shown in Fig. 42.7. Electrons and holes generated in the quantum wells by absorption of photons can escape into the external circuit by tunnelling and/or thermal emission, as illustrated schematically in Fig. 42.7.

The physics of tunnelling in quantum wells is essentially the same as that of α -decay in nuclear physics. The confined particle oscillates within the well and attempts to escape every time it hits the barrier. The escape rate is proportional to the attempt frequency ν_0 and the transmission probability of the barrier. For the simplest case of a rectangular barrier of thickness b , the escape time τ_T is given by:

$$\frac{1}{\tau_T} = \nu_0 \exp(-2\kappa b), \quad (42.23)$$

where κ is the tunnelling decay constant given by (42.10). The factor of 2 in the exponential arises due to the dependence of the transmission probability on $|\psi(z)|^2$. The situation in a biased quantum well is more complicated due to the non-rectangular shape of the barriers. However, (42.23) allows the basic trends to be understood. To obtain fast tunnelling we need thin barriers and small κ . The second requirement is achieved by keeping m_b^* as small as possible and by working with a small confining potential V_0 . The tunnelling rate increases with increasing field, because the average barrier height decreases.

The thermal emission of electrons over a confining potential is an old problem which was originally applied to the heated cathodes in vacuum tubes. It has been shown that the thermal current fits well to the classical Richardson formula:

$$J_E \propto T^{1/2} \exp\left(-\frac{e\Phi}{k_B T}\right), \quad (42.24)$$

with the work function Φ replaced by $[V(F) - E_n]$, $V(F)$ being the height of the barrier that must be overcome at the field strength F [42.15]. The emission rate is dominated by the Boltzmann factor, which represents the probability that the carriers have enough thermal kinetic energy to escape over the top of the barrier. At low fields $V(F) \approx (V_0 - E_n)$, but as the field increases, $V(F)$ decreases as the barriers tilt over. Hence the emission rate (like the tunnelling rate) increases with increasing field. The only material-dependent parameter that enters the Boltzmann factor is the barrier height. Since this is insensitive both to the effective masses and to the barrier

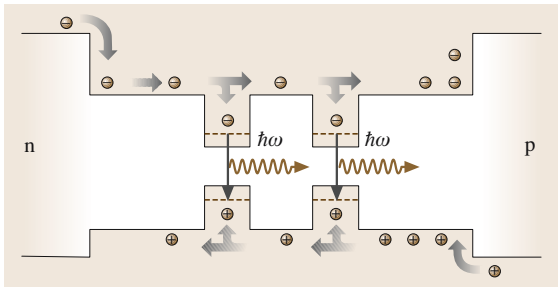


Fig. 42.10 Schematic representation of the drift of injected carriers and their subsequent capture by quantum wells. Light emission occurs when electrons and holes are captured in the same quantum well and then recombine with each other

thickness, the thermal emission rate can dominate over the tunnelling rate in some conditions, especially at room temperature in samples with thick barriers. For example, the fastest escape mechanism in GaAs/Al_{0.3}Ga_{0.7}As QWs at room temperature can be the thermal emission of holes, which have a much smaller barrier to overcome than the electrons [42.16].

Superlattices

The artificial periodicity of superlattice structures gives rise to additional vertical transport effects related to the phenomenon of Bloch oscillations. It is well known that an electron in a periodic structure is expected to oscillate when a DC electric field is applied. This effect has never been observed in natural crystals because the oscillation period – equal to h/eFa , where a is the unit cell size – is much longer than the scattering times of the electrons. In a superlattice, by contrast, the unit cell size is equal to $(d+b)$ (See Fig. 42.3) and the oscillation period can be made much shorter. This allows the electron to perform several oscillations before being scattered. The oscillatory motion of the electrons in a superlattice was first observed by two groups in 1992 [42.17, 18]. The following year, another group directly detected the radiation emitted by the oscillating electron wave packet [42.19]. The subject has since developed greatly, and THz-frequency emission has now been achieved from GaAs/AlGaAs superlattices even at room temperature [42.20].

42.2.6 Carrier Capture and Relaxation

In a QW light-emitting device, the emission occurs after carriers injected from the contacts are transported to the active region and then captured by the QWs. The

capture and subsequent relaxation of the carriers is thus of crucial importance. Let us consider the band-edge profile of a typical QW diode-laser active region, as illustrated in Fig. 42.10. The active region is embedded between larger band-gap cladding layers designed to prevent thermally assisted carrier leakage outside the active region. Electrons and holes are injected from the n- and p-doped cladding layers under forward bias and light emission follows after four distinct processes have taken place: (1) relaxation of carriers from the cladding layers to the confinement barriers (CB); (2) carrier transport across the CB layers, by diffusion and drift; (3) carrier capture into the quantum wells; and (4) carrier relaxation to the fundamental confined levels.

The carrier relaxation to the CB layers occurs mainly by longitudinal optical (LO) phonon emission. The CB layer transport is governed by a classical electron fluid model. The holes are heavier and less mobile than the electrons, and hence the ambipolar transport is dominated by the holes. Carrier nonuniformities, such as carrier pile up at the p-side CB region due to the lower mobility of the holes, are taken into account in the design of the barrier layers. The carrier capture in the QWs is governed by the phonon-scattering-limited carrier mean free path. It is observed experimentally that the capture time oscillates with the QW width. Detailed modelling reveals that this is related to a resonance between the LO phonon energy and the energy difference between the barrier states and the confined states within the well [42.21]. As another design guideline, the QW widths must be larger or at least equal to the phonon-scattering-limited carrier mean free path at the operating temperature in order to speed up the carrier capture. Finally, the relaxation of carriers to the lowest sub-band occurs on a sub-picosecond time scale if the inter-sub-band energy separation is larger than the LO phonon energy. Carrier-carrier scattering can also contribute to an ultrafast thermalisation of carriers, on a femtosecond time scale at the high carrier densities present inside laser diodes. Many of these processes have been studied in detail by ultrafast laser spectroscopy [42.22].

Carrier capture and escape are complementary vertical transport mechanisms in MQW structures. In the design of vertical transport-based MQW devices, one process must often be sped up at the expense of making the other as slow as possible. For example, in order to enhance the performance of QW laser diodes, the carrier confinement capability of the MQW active region must be optimised in terms of minimising the ratio between the carrier capture and escape times [42.23].

42.3 Emitters

42.3.1 Interband Light-Emitting Diodes and Lasers

Quantum wells have found widespread use in light-emitting diode (LED) and laser diode applications for a number of years now. As discussed in Sects. 42.1.2 and 42.2.2, there are three main reasons for this. Firstly, the ability to control the quantum-confinement energy provides an extra degree of freedom to engineer the emission wavelength. Secondly, the change of the density of states and the enhancement of the electron-hole overlap leads to superior performance. Finally, the ability to grow strained layers of high optical quality greatly increases the variety of material combinations that can be employed, thus providing much greater flexibility in the design of the active regions.

Much of the early work concentrated on lattice-matched combinations such as GaAs/AlGaAs on GaAs substrates. GaAs/AlGaAs QW lasers operating around 800 nm has now become industry-standard for applications in laser printers and compact discs. Furthermore, the development of high-power arrays has opened up new applications for pumping solid-state lasers such as Nd : YAG. Other types of lattice-matched combinations can be used to shift the wavelength into the visible spectral region and also further into the infrared. QWs based on the quaternary alloy $(\text{Al}_y\text{Ga}_{1-y})_x\text{In}_{1-x}\text{P}$, are used for red-emitting laser pointers [42.24], while $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ QWs and its variants incorporating Al are used for the important telecommunication wavelengths of 1300 nm and 1550 nm.

The development of strained-layer QW lasers has greatly expanded the range of material combinations that are available. The initial work tended to focus on $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ QWs grown on GaAs substrates. The incorporation of indium into the quantum well shifts the band edge to lower energy, thereby giving emission in the wavelength range 900–1100 nm. An important technological driving force has been the need for powerful sources at 980 nm to pump erbium-doped fibre amplifiers [42.25]. Furthermore, as mentioned in Sect. 42.2.1, the strain alters the band structure and this can have other beneficial effects on the device performance. For example, the compressive strain in the $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ QW system has been exploited in greatly reducing the threshold current density. This property is related to the reduced effective mass of the holes and hence the reduced density of states. An extensive account of the effects of strain on semiconductor

layers and the performance of diode lasers is given in [42.26].

At the other end of the spectral range, a spectacular development has been the $\text{In}_x\text{Ga}_{1-x}\text{N}/\text{GaN}$ QWs grown on sapphire substrates. These highly strained QWs are now routinely used in ultrabright blue and green LEDs, and there is a growing interest in developing high-power LED sources for applications in solid-state lighting [42.27]. Commercial laser diodes operating around 400 nm have been available for several years [42.8], and high-power lasers suited to applications in large-capacity optical disk video recording systems have been reported [42.28]. Lasers operating out to 460 nm have been demonstrated [42.29], and also high-efficiency ultraviolet light-emitting diodes [42.30]. At the same time, much progress has been made in the application of AlGaIn/GaN quantum-well materials in high-power microwave devices [42.31, 32].

A major application of quantum wells is in *vertical-cavity surface-emitting lasers (VCSELs)*. These lasers emit from the top of the semiconductor surface, and have several advantages over the more-conventional edge-emitters: arrays are readily fabricated, which facilitates their manufacture; no facets are required, which avoids complicated processing procedures; the beam is circular, which enhances the coupling efficiency into

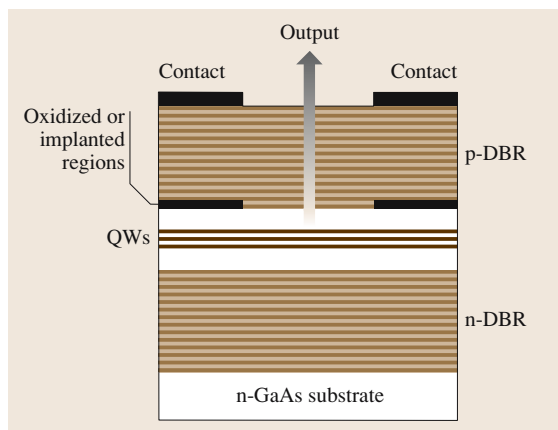


Fig. 42.11 Schematic diagram of a vertical-cavity surface-emitting laser (VCSEL). The quantum wells (QWs) that comprise the gain medium are placed at the centre of the cavity formed between two distributed Bragg reflector (DBR) mirrors. Oxidised or proton-implanted regions provide the lateral confinement for both the current and the optical mode

optical fibres; and their small size leads to very low threshold currents. For these reasons the development of VCSELs has been very rapid, and many local-area fibre networks operating around 850 nm currently employ VCSEL devices. This would not have been possible without the high gain coefficients that are inherent to the QW structures.

Figure 42.11 gives a schematic diagram of a typical GaAs-based VCSEL. The device contains an active QW region inserted between two distributed Bragg reflector (DBR) mirrors consisting of AlGaAs quarter-wave stacks made of alternating high- and low-refractive-index layers. The structure is grown on an n-type GaAs substrate, and the mirrors are appropriately doped n- or p-type to form a p-n junction. Electrons and holes are injected into the active region under forward bias, where they are captured by the QWs and produce gain at the lasing wavelength λ . The quantum wells are contained within a transparent layer of thickness $\lambda/2n_0$, where n_0 is the average refractive index of the active region. The light at the design wavelength is reflected back and forth through the gain medium and adds up constructively, forming a laser resonator. Oxidised or proton-implanted regions provide lateral confinement of both the current and the optical mode. Reviews of the design and properties of VCSELs may be found in [42.34] and [42.35].

The conventional VCSEL structures grown on GaAs substrates operate in the wavelength range 700–1100 nm [42.36]. Some of these structures are lattice-matched, but others – notably the longer-wavelength devices which incorporate strained InGaAs quantum wells – are not. Much work is currently focussed on extending the range of operation to the telecommunication wavelengths of 1300 nm and 1550 nm. Unfortunately, it is hard to grow DBR mirrors with sufficient reflectivity on InP substrates due to the low refractive-index contrast of the materials, and thus progress has been slow. Recent alternative approaches based on GaAs substrates will be mentioned in Sect. 42.6.

Resonator structures such as the VCSEL shown in Fig. 42.11 can be operated below threshold as resonant-cavity LEDs (RCLEDs). The presence of the cavity reduces the emission line width and hence increases the intensity at the peak wavelength [42.37]. Furthermore, the narrower line width leads to an increase in the bandwidth of the fibre communication system due to the reduced chromatic dispersion [42.38]. A review of the progress in RCLEDs is given in [42.39].

42.3.2 Quantum Cascade Lasers

The principles of infrared emission by ISB transitions were described in Sect. 42.2.4. Electrons must first be injected into an upper confined electron level as shown in Fig. 42.9b. Radiative transitions to lower confined states with different parities can then occur. ISB emission is usually very weak, as the radiative transitions have to compete with very rapid nonradiative decay by phonon emission, (Sect. 42.2.6). However, when the electron density in the upper level is large enough, population inversion can occur, giving rapid stimulated emission and subsequent laser operation. This is the operating concept of the *quantum cascade (QC) laser* first demonstrated in 1994 [42.40]. The laser operated at 4.2 μm at temperatures up to 90 K. Although the threshold current for the original device was high, progress in the field had been very rapid. A comprehensive review of the present state of the art is given in [42.33], while a more introductory overview may be found in [42.41].

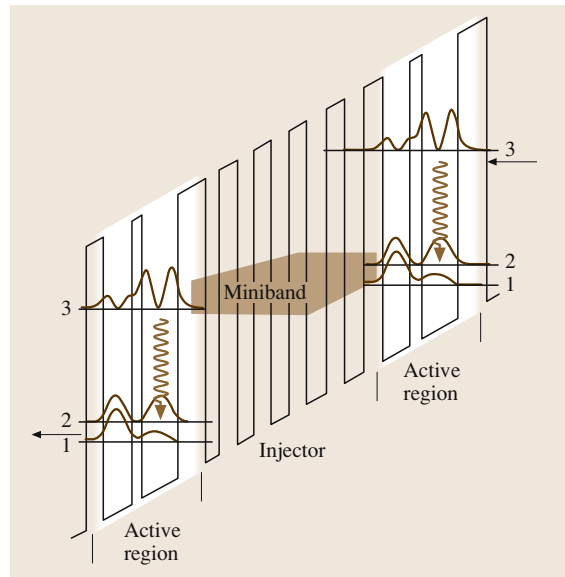


Fig. 42.12 Conduction band diagram for two active regions of an InGaAs/AlInAs quantum cascade laser, together with the intermediate miniband injector region. The levels in each active region are labelled according to their quantum number n , and the corresponding wave function probability densities are indicated. Laser transitions are indicated by the wavy arrows, while electron tunnelling processes are indicated by the straight arrows. (After [42.33], © 2001 IOP)

The quantum-well structures used in QC lasers are very complicated, and often contain hundreds of different layers. Figure 42.12 illustrates a relatively simple design based on lattice-matched $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}/\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ quantum wells grown on an InP substrate. The diagram shows two active regions and the miniband injector region that separates them. A typical operational laser might contain 20–30 such repeat units. The population inversion is achieved by resonant tunnelling between the $n = 1$ ground state of one active region and the $n = 3$ upper laser level of the next one. The basic principles of this process were enunciated as early as 1971 [42.42], but it took more than 20 years to demonstrate the ideas in the laboratory. The active regions contain asymmetric coupled quantum wells, and the laser transition takes place between the $n = 3$ and $n = 2$ states of the coupled system. The separation of the $n = 2$ and $n = 1$ levels is carefully designed to coincide with the LO-phonon energy, so that very rapid relaxation to the ground state oc-

curs and the system behaves as a four-level laser. This latter point is crucial, since the lifetime of the upper laser level is very short (typically ≈ 1 ps), and population inversion is only possible when the lifetime of the lower laser level is shorter than that of the upper one. The lasing wavelength can be varied by detailed design of the coupled QW active region. The transition energy for the design shown in Fig. 42.12 is 0.207 eV, giving emission at $6.0\text{ }\mu\text{m}$. Further details may be found in [42.33].

A very interesting recent development has been the demonstration of a QC laser operating in the far-infrared spectral region at $67\text{ }\mu\text{m}$ [42.43]. Previous work in this spectral region had been hampered by high losses due to free-carrier absorption and the difficulties involved in designing the optical waveguides. The device operated up to 50 K and delivered 2 mW. These long-wavelength devices are required for applications in the THz frequency range that bridges between long-wavelength optics and high-frequency electronics.

42.4 Detectors

Photodetectors for the visible and near-infrared spectral regions are generally made from bulk silicon or III–V alloys such as GaInAs. Since these devices work very well, the main application for QW photodetectors is in the infrared spectral region and for especially demanding applications such as avalanche photodiodes and solar cells. These three applications are discussed separately below, starting with solar cells.

42.4.1 Solar Cells

The power generated by a solar cell is determined by the product of the photocurrent and the voltage across the diode. In conventional solar cells, both of these parameters are determined by the band gap of the semiconductor used. Large photocurrents are favoured by narrow-gap materials, because semiconductors only absorb photons with energies greater than the band gap, and narrow-gap materials therefore absorb a larger fraction of the solar spectrum. However, the largest open-circuit voltage that can be generated in a p–n device is the built-in voltage which increases with the band gap of the semiconductor. Quantum-well devices can give better performance than their bulk counterparts because they permit separate optimisation of the current- and voltage-generating factors [42.44]. This is because the built-in voltage is

primarily determined by the band gap of the barrier regions, whereas the absorption edge is determined by the band gap of the quantum wells. The drawback in using quantum wells is that it is difficult to maintain high photocurrent quantum efficiency in the low-field forward-bias operating conditions in solar cells.

Recent work in this field has explored the added benefits of the versatility of the design of the QW active region [42.45] and also the possibility of using strained QWs. In the latter case, a tradeoff arises between the increase in both the absorption and the number of interface dislocations (which act as carrier traps) with the number of QWs. A way round this compromise is to use strain balance. An example is the case of $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}_{0.94}\text{P}_{0.06}$ QW solar cells grown on GaAs substrates, in which the compressive strain of the InGaAs QWs is compensated with the tensile-strained $\text{GaAs}_{0.94}\text{P}_{0.06}$ barriers, such that the overall active region could be successfully lattice-matched to the substrate [42.46].

42.4.2 Avalanche Photodiodes

Avalanche photodiodes (APDs) are the detectors of choice for many applications in telecommunications and single-photon counting. The avalanche multiplica-

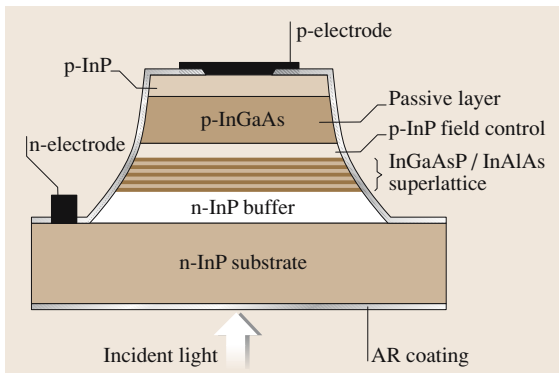


Fig. 42.13 Schematic representation of an InGaAs/InP/InGaAsP/InAlAs superlattice avalanche photodiode (SL-APD). Light is absorbed in the bulk InGaAs layer and the resulting photocurrent is multiplied by the avalanche process in the InGaAsP/InAlAs superlattice region. The spatial periodicity of the superlattice reduces the dark current

tion mechanism plays a critical role in determining the photodetection gain, the noise and the gain-bandwidth product. Commercially available III–V semiconductor APDs are typically engineered with different band-gap materials in the absorption and multiplication regions [42.47]. The absorption layer has a relatively narrow band gap (such as $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$) to allow for larger absorption, whereas the multiplication region has a wider band gap (such as InP or $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$) to reduce the dark current at the high electric fields required. It has been demonstrated that the dark current can be further reduced by incorporating adequately designed superlattices into the multiplication layer to form a *superlattice avalanche photodiode* (SL-APD) as shown in Fig. 42.13 [42.48]. Essentially, the miniband formation (Fig. 42.3) in the SL corresponds to a larger effective band gap and thus reduces the probability of band-to-band tunnelling at the high electric fields required for carrier impact ionisation and subsequent carrier multiplication.

It has been proposed that the excess noise factor can be reduced further by designing multiplication layers with SL [42.49] or *staircase* [42.50] structures to enhance the ionisation rate of one carrier type relative to the other due to their different band-edge discontinuities [42.51]. However, this proposal is still under active theoretical and experimental scrutiny, because most of the electrons populate the higher energy satellite valleys (X and L) at the high electric fields required for avalanche gain [42.52]. Recent Monte Carlo

simulations [42.53], backed up by experimental evidence [42.54], have shown that the ratio of the electron and hole ionisation rates, commonly used as a figure of merit in bulk multiplication layers [42.55], is not critically affected by the band-gap discontinuities. In fact, these recent models emphasise that it is the spatial modulation of the impact ionisation probability, associated with the periodic band-gap discontinuity, which leads to a reduction in the multiplication noise.

42.4.3 Inter-Sub-Band Detectors

The principles of infrared absorption by ISB transitions were described in Sect. 42.2.4. Infrared detectors are required for applications in defence, night vision, astronomy, thermal mapping, gas-sensing, and pollution monitoring. *Quantum-well inter-sub-band photodetectors* (QWIPs) are designed so that the energy separation of the confined levels is matched to the chosen wavelength. A major advantage of QWIPs over the conventional approach employing narrow-gap semiconductors is the use of mature GaAs-based technologies. Furthermore, the detection efficiency should in principle be high due to the large oscillator strength that follows from the parallel curvature of the in-plane dispersions for states within the same bands [42.56]. However, technical challenges arise from the requirement that the electric field of the light must be polarised along the growth (z) direction. This means that QWIPs do not work at normal incidence, unless special steps are taken to introduce a component of the light field along the z -direction. Various approaches have been taken for optimum light coupling, such as using bevelled edges, gratings or random reflectors [42.57].

Despite their promising characteristics, QWIPs have yet to be commercialised. The main issue is the high dark current at higher operating temperatures. The dark current is governed by the thermionic emission of ground-state electrons directly out of the QW above 45 K [42.58]. Overcoming such technical difficulties has made possible the demonstration of long-wavelength large-format focal-plane array cameras based on ISB transitions [42.59].

42.4.4 Unipolar Avalanche Photodiodes

The combination of resonant inter-sub-band (ISB) photodetection (Sect. 42.4.3) and avalanche multiplication has been studied for exploiting the possibility of designing a unipolar avalanche photodiode (UAPD) [42.60]. UAPDs rely on impact ionisation involving only one

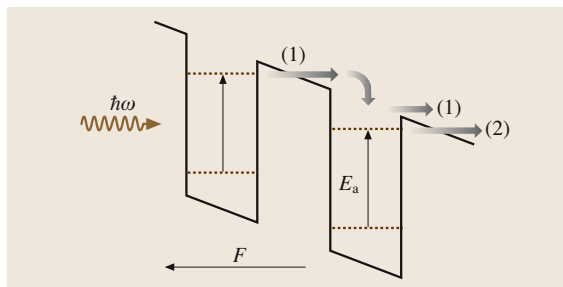


Fig. 42.14 Unipolar carrier multiplication in a multiple QW structure at field strength F : (1) is the primary electron resulting from photodetection, while (2) is the secondary electron resulting from the impact ionisation of the QW by the primary electron

type of carrier in order to achieve gain in photoconductive detectors for mid- and far-infrared light. The unipolar impact ionisation occurs when the kinetic energy of the primary carrier exceeds an activation energy E_a defined as the transition energy between the QW ground state and the QW top state (Fig. 42.14). A single QW then releases an extra electron each time it is subject to an impact with an incoming electron.

The impact ionisation probability for this process is given by the product between the carrier capture probability and the carrier escape probability under

a mechanism of carrier–carrier scattering in the QW. The subsequent electron transport towards further multiplication events occurs through a sequence of escape, drift and kinetic energy gain under the applied electric field, relaxation, capture and QW ionisation. Ultimately, the multiplication gain in an UAPD is governed by the QW capture probability, the number of QWs and the field uniformity over the QW sequence. The unipolar nature of the multiplication process must be preserved in order to avoid field nonuniformities stemming from spatial-charge variation caused by bipolar carrier transport across the multiplication region.

Interest in a purely unipolar multiplication mechanism was originally motivated by the possibility of reduced noise in comparison to bipolar APDs (Sect. 42.4.2), where band-to-band transitions lead to gain fluctuations manifested as excess noise [42.55]. For this purpose, the QWs in an UAPD structure are typically tailored such that the inter-sub-band activation energy E_a is smaller than the inter-band impact ionisation activation energy that would be responsible for bipolar avalanche multiplication. However, recent studies have shown that unipolar avalanche multiplication is also accompanied by an excess noise factor, such that the noise gain exceeds the photoconductive gain [42.61], thus limiting the practical applications of UAPDs.

42.5 Modulators

In Sect. 42.2.3 we noted that the optical properties of quantum-well diodes are strongly modified by the application of voltages through the quantum-confined Stark effect (QCSE). Referring to Fig. 42.8, we see that at wavelengths below the heavy-hole exciton at 0 V, the absorption increases as the voltage is applied, which provides a mechanism for the modulation of light. For example, the amount of light transmitted at wavelengths close to the band edge would change with the voltage applied. Moreover, since changes of absorption are accompanied by changes of the refractive index through the Kramers–Kronig relationship, it is possible to make QCSE phase modulators as well [42.62]. In addition to the standard GaAs/AlGaAs devices operating around 800 nm, QCSE modulators have been demonstrated in several other material systems, such as GaInAs-based structures for the important telecommunications wavelength at 1.5 μm [42.63].

The operation of GaAs-based QCSE transmission modulators at normal incidence is hampered

by the fact that the substrates are opaque at their operating wavelength. One way round this problem

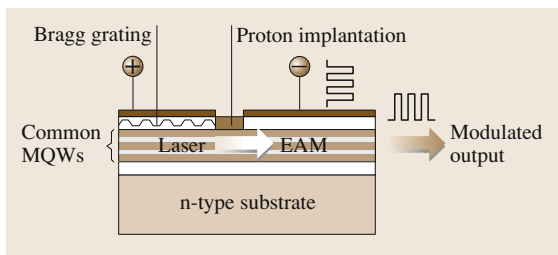


Fig. 42.15 Schematic diagram of an integrated quantum-well waveguide electroabsorption modulator (EAM) and distributed feedback (DFB) laser. The laser is forward-biased while the EAM is reverse-biased. The p-contacts of the two electrically independent devices are separated by proton implantation. The light emitted by the laser is guided through the EAM region, resulting in a modulated output beam when data pulses are applied to the EAM

is to etch the substrate away, but this is a difficult process, and a much better solution is to include a mirror underneath the quantum wells so that the modulated light does not have to pass through the substrate [42.64]. In many practical applications, however, mostly involving the integration of **QCSE** modulators with **MQW** light emitters on a common substrate, the waveguide geometry is the configuration of choice [42.65]. In this architecture, the light beam propagates along the waveguide from the emitter to the *electroabsorption modulator* (EAM) region, as shown in Fig. 42.15. The **QCSE** modulator transmits the incoming laser light when no voltage is applied and absorbs the beam when the **MQW** stack is suitably biased.

The most successful commercial impact of **QCSE** modulators has been in the integration of EAMs with distributed feedback (**DFB**) or distributed Bragg reflector (**DBR**) **MQW** diode lasers in waveguide configurations, as shown in Fig. 42.15. These devices have been used for optical coding in the C-band (1525–1565 nm), at 10 Gb/s or higher data transmission speeds [42.66]. The

combination of a continuous laser and a high-speed modulator offers better control of the phase chirp of the pulses than direct modulation of the laser output itself [42.67]. In particular, the chirp factor is expected to become negative if the photogenerated carriers can be swept out fast enough in the EAM, which is desirable for long-distance data transmission through optical fibers [42.68].

A promising step toward the merger between band-gap-engineered semiconductors and mature very-large-scale integration (VLSI) silicon architectures has been achieved when III–V semiconductor **QCSE** modulator structures have been integrated with state-of-the-art silicon complementary metal–oxide–semiconductor (**CMOS**) circuitry [42.69]. Through this hybrid technology, thousands of optical inputs and outputs could be provided to circuitry capable of very complex information processing. The idea of using light beams to replace wires in telecommunications and digital computer systems has thus become an attractive technological avenue in spite of various challenges implied [42.70].

42.6 Future Directions

The subject of quantum-confined semiconductor structures moves very rapidly and it is difficult to see far into the future. Some ideas have moved very quickly from the research labs into the commercial sector (e.g. **VCSELs**), while others (e.g. quantum cascade lasers) have taken many years to come to fruition. We thus restrict ourselves here to a few comments based on active research fields at the time of writing.

One idea that is being explored in detail is the effects of lower dimensionality in quantum-wire and quantum-dot structures (Table 42.1). Laser operation from one-dimensional (1-D) GaAs quantum wires was first demonstrated in 1989 [42.72], but subsequent progress has been relatively slow due to the difficulty in making the structures. By contrast, there has been an explosion of interest in zero-dimensional (0-D) structures following the discovery that quantum dots can form spontaneously during **MBE** growth in the Stranski–Krastanow regime. A comprehensive review of this subject is given in [42.73].

Figure 42.16 shows an electron microscope image of an InAs quantum dot grown on a GaAs crystal by the Stranski–Krastanow technique. The dots are formed because of the very large mismatch between the lattice constants of the InAs and the underlying GaAs.

The strain that would be produced in a uniform layer is so large that it is energetically favourable to form small clusters. This then leads to the formation of islands of InAs with nanoscale dimensions, which can then be encapsulated within an optoelectronic structure by overgrowth of further GaAs layers.

The ability to grow quantum-dot structures directly by **MBE** has led to very rapid progress in the deployment of quantum dots in a variety of applications. It remains unclear at present whether quantum dots really lead to superior laser performance [42.74]. The intrinsic

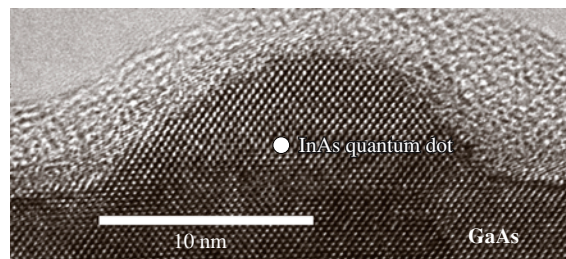


Fig. 42.16 Transmission electron microscope image of an uncapped InAs quantum dot grown on GaAs by the Stranski–Krastanow technique. (After [42.71], © 2000 APS)

sic gain of the dots is higher than that of a quantum well [42.75], and the threshold current is less sensitive to temperature [42.76]. However, the volume of the gain medium is necessarily rather small, and the benefits of the lower dimensionality cannot be exploited to the full. At present, one of the most promising applications for quantum dots is in long-wavelength lasers [42.77]. As mentioned in Sect. 42.3.1, the production of VCSELs at 1300 nm and 1550 nm has proven to be difficult using conventional InP-based QW structures due to the low refractive-index contrast of the materials that form the DBR mirrors. The use of InAs/GaAs quantum dots as the active region circumvents this problem and allows the benefits of mature GaAs-based VCSEL technology.

Another very exciting potential application for quantum dots is in quantum information processing. High-efficiency single-photon sources are required for quantum cryptography and also quantum computation using linear optics. Several groups have demonstrated

single photon emission after excitation of individual InAs quantum dots (see e.g. [42.78, 79]), and one group has demonstrated an electrically driven single-photon LED [42.80]. After these proofs of principle, the challenge now lies ahead to establish the quantum-dot sources in working quantum information-processing systems.

At the same time as exploring the effects of lower dimensionality, many other groups are working on new QW materials. One of the most promising recent developments is the dilute nitride system for applications in long-wavelength VCSELs and solar cells [42.81]. It has been found that the inclusion of a small fraction of nitrogen into GaAs leads to a sharp decrease in the band gap due to very strong band-bowing effects. This then allows the growth of InGaAsN structures that emit at 1300 nm on GaAs substrates [42.77, 82]. The field is developing very rapidly, with 1300-nm VCSELs and 1500-nm edge emitters already demonstrated [42.83, 84].

42.7 Conclusions

Semiconductor quantum wells are excellent examples of quantum mechanics in action. The reduced dimensionality has led to major advances in both the understanding of 2-D physics and the applied science of optoelectronics. In some cases, QWs have enhanced the performance of conventional devices (e.g. LEDs and edge-emitting lasers), and in others, they have led to radically new devices (e.g. VCSELs, quantum cascade lasers, QCSE

modulators). At present, the main commercial use for QW optoelectronic devices is in LEDs, laser diodes and QCSE modulators. It remains to be seen whether some of the other devices described here (QW solar cells, SL-APDs, QWIPs) will come to commercial fruition, and whether systems of lower dimensionality will eventually replace QWs in the same way that QWs have replaced bulk devices.

References

- 42.1 L. Esaki, R. Tsu: IBM J. Res. Develop. **14**, 61–5 (1970)
- 42.2 G. Bastard: *Wave Mechanics Applied to Semiconductor Heterostructures* (Wiley, New York 1988)
- 42.3 M. Jaros: *Physics and Applications of Semiconductor Microstructures* (Clarendon, Oxford 1989)
- 42.4 C. Weisbuch, B. Vinter: *Quantum Semiconductor Structures* (Academic, San Diego 1991)
- 42.5 S. O. Kasap: *Optoelectronics and Photonics: Principles and Practices* (Prentice Hall, Upper Saddle River 2001)
- 42.6 M. J. Kelly: *Low-Dimensional Semiconductors* (Clarendon, Oxford 1995)
- 42.7 Paul J. Dean: III–V Compound Semiconductors. In: *Electroluminescence*, ed. by J. I. Pankove (Springer, Berlin, Heidelberg 1977) pp. 63–132
- 42.8 S. Nakamura, S. Pearton, G. Fasol: *The Blue Laser Diode*, 2nd edn. (Springer, Berlin, Heidelberg 2000)
- 42.9 S. Gasiorowicz: *Quantum Physics*, 2nd edn. (Wiley, New York 1996)
- 42.10 E. P. O'Reilly: Semicond. Sci. Technol. **4**, 121–137 (1989)
- 42.11 M. Fox: *Optical Properties of Solids* (Clarendon, Oxford 2001)
- 42.12 M. Shinada, S. Sugano: J. Phys. Soc. Jpn. **21**, 1936–46 (1966)
- 42.13 D. A. B. Miller, D. S. Chemla, D. J. Eilenberger, P. W. Smith, A. C. Gossard, W. T. Tsang: Appl. Phys. Lett. **41**, 679–81 (1982)
- 42.14 A. M. Fox, D. A. B. Miller, G. Livescu, J. E. Cunningham, W. Y. Jan: IEEE J. Quantum Electron. **27**, 2281–95 (1991)

- 42.15 H. Schneider, K. von Klitzing: Phys. Rev. B **38**, 6160–5 (1988)
- 42.16 A. M. Fox, R. G. Ispasoiu, C. T. Foxon, J. E. Cunningham, W. Y. Jan: Appl. Phys. Lett. **63**, 2917–9 (1993)
- 42.17 J. Feldmann, K. Leo, J. Shah, D. A. B. Miller, J. E. Cunningham, T. Meier, G. von Plessen, A. Schulze, P. Thomas, S. Schmitt-Rink: Phys. Rev. B **46**, 7252–5 (1992)
- 42.18 K. Leo, P. Haring Bolivar, F. Brüggemann, R. Schwedler, K. Köhler: Solid State Commun. **84**, 943–6 (1992)
- 42.19 C. Waschke, H. G. Roskos, R. Schwedler, K. Leo, H. Kurz, K. Köhler: Phys. Rev. Lett. **70**, 3319–22 (1993)
- 42.20 Y. Shimada, K. Hirakawa, S.-W. Lee: Appl. Phys. Lett. **81**, 1642–4 (2002)
- 42.21 P. W. M. Blom, C. Smit, J. E. M. Haverkort, J. H. Wolter: Phys. Rev. B **47**, 2072–2081 (1993)
- 42.22 J. Shah: *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, 2nd edn. (Springer, Berlin, Heidelberg 1999)
- 42.23 R. G. Ispasoiu, A. M. Fox, D. Botez: IEEE J. Quantum Electron. **36**, 858–63 (2000)
- 42.24 P. Blood: Visible-emitting quantum well lasers. In: *Semiconductor Quantum Optoelectronics*, ed. by A. Miller, M. Ebrahimzadeh, D. M. Finlayson (Institute of Physics, Bristol 1999) pp. 193–211
- 42.25 N. Chand, S. N. G. Chu, N. K. Dutta, J. Lopata, M. Geva, A. V. Syrbu, A. Z. Mereutza, V. P. Yakovlev: IEEE J. Quantum Electron. **30**, 424–40 (1994)
- 42.26 E. P. O'Reilly, A. R. Adams: IEEE J. Quantum Electron. **30**, 366–79 (1994)
- 42.27 M. R. Krames, J. Bhat, D. Collins, N. F. Gargner, W. Götz, C. H. Lowery, M. Ludowise, P. S. Martin, G. Mueller, R. Mueller-Mach, S. Rudaz, D. A. Steigerwald, S. A. Stockman, J. J. Wierer: Phys. Stat. Sol. A **192**, 237–245 (2002)
- 42.28 M. Ikeda, S. Uchida: Phys. Stat. Sol. A **194**, 407–13 (2002)
- 42.29 S. Nagahama, T. Yanamoto, M. Sano, T. Mukai: Phys. Stat. Sol. A **194**, 423–7 (2002)
- 42.30 S. Kamiyama, M. Iwaya, H. Amano, I. Akasaki: Phys. Stat. Sol. A **194**, 393–8 (2002)
- 42.31 L. F. Eastman, V. Tilak, V. Kaper, J. Smart, R. Thompson, B. Green, J. R. Shealy, T. Prunty: Phys. Stat. Sol. A **194**, 433–8 (2002)
- 42.32 W. S. Tan, P. A. Houston, P. J. Parbrook, G. Hill, R. J. Airey: J. Phys. D: Appl. Phys. **35**, 595–8 (2002)
- 42.33 C. Gmachl, F. Capasso, D. L. Sivco, A. Y. Cho: Rep. Prog. Phys. **64**, 1533–1601 (2001)
- 42.34 K. J. Eberling: Analysis of vertical cavity surface emitting laser diodes (VCSEL). In: *Semiconductor Quantum Optoelectronics*, ed. by A. Miller, M. Ebrahimzadeh, D. M. Finlayson (Institute of Physics, Bristol 1999) pp. 295–338
- 42.35 M. SanMiguel: Polarisation properties of vertical cavity surface emitting lasers. In: *Semiconductor Quantum Optoelectronics*, ed. by A. Miller, M. Ebrahimzadeh, D. M. Finlayson (Institute of Physics, Bristol 1999) pp. 339–366
- 42.36 O. Blum Spahn: Materials issues for vertical cavity surface emitting lasers (VCSEL) and edge emitting lasers (EEL). In: *Semiconductor Quantum Optoelectronics*, ed. by A. Miller, M. Ebrahimzadeh, D. M. Finlayson (Institute of Physics, Bristol 1999) pp. 265–94
- 42.37 E. F. Schubert, Y.-H. Wang, A. Y. Cho, L.-W. Tu, G. J. Zydzik: Appl. Phys. Lett. **60**, 921–3 (1992)
- 42.38 N. E. Hunt, E. F. Schubert, R. F. Kopf, D. L. Sivco, A. Y. Cho, G. J. Zydzik: Appl. Phys. Lett. **63**, 2600–2 (1993)
- 42.39 R. Baets: Micro-cavity light emitting diodes. In: *Semiconductor Quantum Optoelectronics*, ed. by A. Miller, M. Ebrahimzadeh, D. M. Finlayson (Institute of Physics, Bristol 1999) pp. 213–64
- 42.40 J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, A. Y. Cho: Science **264**, 553–6 (1994)
- 42.41 F. Capasso, C. Gmachl, D. L. Sivco, A. Y. Cho: Phys. Today **55**(5), 34–40 (2002)
- 42.42 R. A. Kazarinov: Sov. Phys. Semicond. **5**, 707–9 (1971)
- 42.43 R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, F. Rossi: Nature **417**, 156–9 (2002)
- 42.44 K. Barnham, I. Ballard, J. Barnes, J. Connolly, P. Griffin, B. Klufftinger, J. Nelson, E. Tsui, A. Zachariou: Appl. Surf. Sci. **113/114**, 722–733 (1997)
- 42.45 R. H. Morf: Physica E **14**, 78–83 (2002)
- 42.46 N. J. Ekins-Daukes, K. W. J. Barnham, J. P. Connolly, J. S. Roberts, J. C. Clark, G. Hill, M. Mazzer: Appl. Phys. Lett. **75**, 4195–5197 (1999)
- 42.47 J. Wei, J. C. Dries, H. Wang, M. L. Lange, G. H. Olsen, S. R. Forrest: IEEE Photon. Technol. Lett. **14**, 977–9 (2002)
- 42.48 A. Suzuki, A. Yamada, T. Yokotsuka, K. Idota, Y. Ohiki: Jpn. J. Appl. Phys. **41**, 1182–5 (2002)
- 42.49 F. Capasso, W. T. Tsang, A. L. Hutchinson, G. F. Williams: Appl. Phys. Lett. **40**, 38–40 (1982)
- 42.50 G. Ripamonti, F. Capasso, A. L. Hutchinson, D. J. Muehlner, J. F. Walker, R. J. Malek: Nucl. Instrum. Meth. Phys. Res. A **288**, 99–103 (1990)
- 42.51 R. Chin, N. Holonyak, G. E. Stillman, J. Y. Tang, K. Hess: Electron. Lett. **16**, 467–9 (1980)
- 42.52 C. K. Chia, J. P. R. David, G. J. Rees, P. N. Robson, S. A. Plimmer, R. Grey: Appl. Phys. Lett. **71**, 3877–9 (1997)
- 42.53 F. Ma, X. Li, S. Wang, K. A. Anselm, X. G. Zheng, A. L. Holmes, J. C. Campbell: J. Appl. Phys. **92**, 4791–5 (2002)
- 42.54 P. Yuan, S. Wang, X. Sun, X. G. Zheng, A. L. Holmes, J. C. Campbell: IEEE Photon. Technol. Lett. **12**, 1370–2 (2000)

- 42.55 M. A. Saleh, M. M. Hayat, P. P. Sotirelis, A. L. Holmes, J. C. Campbell, B. E. A. Saleh, M. C. Teich: *IEEE Trans. Electron. Dev.* **48**, 2722–31 (2001)
- 42.56 L. C. West, S. J. Eglash: *Appl. Phys. Lett.* **46**, 1156–8 (1985)
- 42.57 S. D. Gunapala, G. Sarusi, J. S. Park, T. Lin, B. F. Levine: *Phys. World* **7**(10), 35–40 (1994)
- 42.58 S. D. Gunapala, S. V. Bandara: Quantum well infrared photodetector (QWIP) focal plane arrays. In: *Semiconductors and Semimetals*, Vol. 62, ed. by M. C. Liu, F. Capasso (Academic, New York 1999) pp. 197–282
- 42.59 S. D. Gunapala, S. V. Bandara, J. K. Liu, E. M. Luong, N. Stetson, C. A. Shott, J. J. Block, S. B. Rafol, J. M. Mumolo, M. J. McKelvey: *IEEE Trans. Electron. Dev.* **47**, 326–332 (2000)
- 42.60 B. F. Levine, K. K. Choi, C. G. Bethea, J. Walker, R. J. Malik: *Appl. Phys. Lett.* **51**, 934–6 (1987)
- 42.61 H. Schneider: *Appl. Phys. Lett.* **82**, 4376–8 (2003)
- 42.62 J. S. Weiner, D. A. B. Miller, D. S. Chemla: *Appl. Phys. Lett.* **50**, 842–4 (1987)
- 42.63 R. W. Martin, S. L. Wong, R. J. Nicholas, K. Satzke, M. Gibbons, E. J. Thrush: *Semicond. Sci. Technol.* **8**, 1173–8 (1993)
- 42.64 G. D. Boyd, D. A. B. Miller, D. S. Chemla, S. L. McCall, A. C. Gossard, J. H. English: *Appl. Phys. Lett.* **50**, 1119–21 (1987)
- 42.65 A. Ramdane, F. Devaux, N. Souli, D. Delprat, A. Ougazzaden: *IEEE J. Quantum Electron.* **2**, 326–35 (1996)
- 42.66 T. Ido, S. Tanaka, M. Suzuki, M. Koizumi, H. Sano, H. Inoue: *J. Lightwave Technol.* **14**, 2026–33 (1996)
- 42.67 Y. Miyazaki, H. Tada, T. Aoyagi, T. Nishimura, Y. Mitsui: *IEEE J. Quantum Electron.* **38**, 1075–80 (2002)
- 42.68 G. Agrawal: *Fiber Optic Communication Systems* (Wiley, New York 1993)
- 42.69 K. W. Goosen, J. A. Walker, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. D. Bacon, D. Dahringer, L. M. F. Chirovsky, A. L. Lentine, D. A. B. Miller: *IEEE Photon. Technol. Lett.* **7**, 360–2 (1995)
- 42.70 D. A. B. Miller: *IEEE J. Sel. Top. Quantum Electron.* **6**, 1312–7 (2000)
- 42.71 P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. J. Finley, J. A. Barker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksym, M. Hopkinson, M. Al-Khafaji, J. P. R. David, A. G. Cullis, G. Hill, J. C. Clark: *Phys. Rev. Lett.* **84**, 733–6 (2000)
- 42.72 E. Kapon, D. M. Hwang, R. Bhat: *Phys. Rev. Lett.* **63**, 430–3 (1989)
- 42.73 D. Bimberg, M. Grundmann, Nikolai N. Ledentsov: *Quantum Dot Heterostructures* (Wiley, Chichester 1998)
- 42.74 M. Grundmann: *Physica E* **5**, 167–84 (2000)
- 42.75 M. Asada, Y. Miyamoto, Y. Suematsu: *IEEE J. Quantum Electron.* **22**, 1915–21 (1986)
- 42.76 Y. Arakawa, H. Sakaki: *Appl. Phys. Lett.* **40**, 939–41 (1982)
- 42.77 V. M. Ustinov, A. E. Zhukov: *Semicond. Sci. Technol.* **15**, R41–R54 (2000)
- 42.78 P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, A. Imamoglu: *Science* **290**, 2282–5 (2000)
- 42.79 C. Santori, M. Pelton, G. Solomon, Y. Dale, Y. Yamamoto: *Phys. Rev. Lett.* **86**, 1502–5 (2001)
- 42.80 Z. Yuan, B. E. Kardynal, R. M. Stevenson, A. J. Shields, C. J. Lobo, K. Cooper, N. S. Beattie, D. A. Ritchie, M. Pepper: *Science* **295**, 102–5 (2002)
- 42.81 A. Mascarenhas, Y. Zhang: *Current Opinion Solid State Mater. Sci.* **5**, 253–9 (2001)
- 42.82 A. Yu. Egorov, D. Bernklau, B. Borchert, S. Illek, D. Livshits, A. Rucki, M. Schuster, A. Kaschner, A. Hoffmann, Gh. Dumitras, M. C. Amann, H. Riechert: *J. Cryst. Growth* **227–8**, 545–552 (2001)
- 42.83 G. Steinle, H. Riechert, A. Yu. Egorov: *Electron. Lett.* **37**, 93–5 (2001)
- 42.84 D. Gollub, M. Fischer, A. Forchel: *Electron. Lett.* **38**, 1183–4 (2002)

Glasses for Photonic Integration

Inorganic glasses are the workhorse materials of optics and photonics. In addition to offering a range of transparency windows, glasses provide flexibility of processing for the realization of fibers, films, and shaped optical elements. Traditionally, the main role of glass has been as a passive material. However, a significant attribute of glasses is their ability to incorporate dopants such as nanoparticles or active ions. Hence, glasses promise to play an increasingly important role in active photonics, as laser, amplification, switching, and nonlinear media.

For photonic integration, many of the attributes of glasses are particularly compelling. Glasses allow numerous options for thin film deposition and integration on arbitrary platforms. The possibility of controlling the viscosity of a glass during processing can be exploited in the realization of extremely low loss microphotonic waveguides, photonic crystals, and microcavities. The metastable nature of glass can enable the direct patterning of photonic elements by energetic beams.

This chapter provides an overview of these unique properties of glasses, from the perspectives of the technology options they afford and the

43.1 Main Attributes of Glasses as Photonic Materials	1042
43.1.1 The Glass Transition as Enabler ...	1043
43.1.2 Metastability	1046
43.1.3 Glass as Host Material	1049
43.2 Glasses for Integrated Optics	1050
43.2.1 Low Index Glassy Films	1050
43.2.2 Medium Index Glassy Films	1051
43.2.3 High Index Glassy Films	1051
43.3 Laser Glasses for Integrated Light Sources	1053
43.3.1 Advantages of Glass-based Light Sources	1053
43.3.2 Alternative Glass Hosts	1054
43.3.3 Progress Towards Integrated Light Sources in Glass	1056
43.4 Summary	1057
References	1059

practical limitations they present. Further, an overview is provided of the main families of glassy inorganic films studied for integrated optics. Finally, the main features of rare earth doped glasses are reviewed, with an emphasis on their potential for implementation of compact integrated light sources and amplifiers.

Inorganic glasses have played a central role in optical science and technology, and more generally within the electrical and electronic engineering disciplines [43.1]. Amongst optical materials, the unique advantages of glasses are well known [43.2, 3]. Glasses can be worked relatively easily into various forms, such as bulk lenses, fibers, and thin films. This is dramatically illustrated by the modern technology for the manufacture of low-loss silica fibers, in which a hair-thin glass fiber (with tightly controlled geometrical and material properties) is drawn from a heated glass preform at rates as high as 20 m/s or greater [43.4]. Glass can be manufactured with excellent homogeneity and without grain boundaries, so that scattering of light is acceptably low. While perfect crystals typically exhibit even lower lev-

els of scattering, they are difficult to realize in large sizes and with arbitrary shapes. Optical losses in polycrystalline materials, on the other hand, are generally excessive due to scattering from grain boundaries. Finally, the standard glass compositions (oxide, halide, and chalcogenide glasses and hybrids of those) provide transparency windows from the UV to the mid-infrared (see Fig. 43.1). It has been estimated [43.2] that more than 90% of optical components are based on glasses. Given the massive worldwide installation of telecommunications fibers since the 1980s [43.5], and depending on the definition assigned to the term 'component', this statistic might actually be closer to 100%.

Befitting their central role in optical technology, there are numerous excellent reviews covering

the optical properties and photonic functionalities of glasses [43.3, 6]. Further, one of the distinguishing attributes of glasses is their flexibility of composition. Hundreds of commercial glasses are available, and many more have been studied in research laboratories and reported in the academic literature. Thus, it is not possible to capture the full range of glassy materials (even restricted to inorganic glasses) within a short review paper. The recent monograph by *Yamane and Asahara* [43.6] provides a comprehensive and highly recommended treatment of glass technology and its application to photonics. The present contribution has relatively modest aims:

1. With a focus specifically on *integrated* photonics, the main advantages of glasses are discussed and the most studied glass systems are reviewed in brief. Integrated photonics (integrated optics) refers to guided wave photonic devices fabricated on planar platforms. Glass substrates (with waveguides defined by ion diffusion) have played an important role in integrated optics, but the advantages of glasses (compared to crystalline materials such as III-V semiconductors or lithium niobate) are less compelling in that case. For this reason, emphasis is placed on glassy thin films that can enable photonic functionality on other substrates. In practice, substrates of interest might be printed circuit boards or semiconductor wafers.
2. For the most part, the discussion is restricted to inorganic glasses. Some reference is made to organic glasses, since their properties and advantages parallel those of inorganic glasses in many respects, and because certain practical aspects of organic glasses have received greater attention in the literature. Good reviews covering polymer materials for integrated photonics are available [43.8].
3. To date, the most important inorganic glass system for integrated photonics is the silica on silicon planar lightwave circuit (PLC) platform. Several recent papers [43.9, 10] provide an overview of that technology, so only brief mention is made here. Rather, the goal is to provide some insight on advantages that

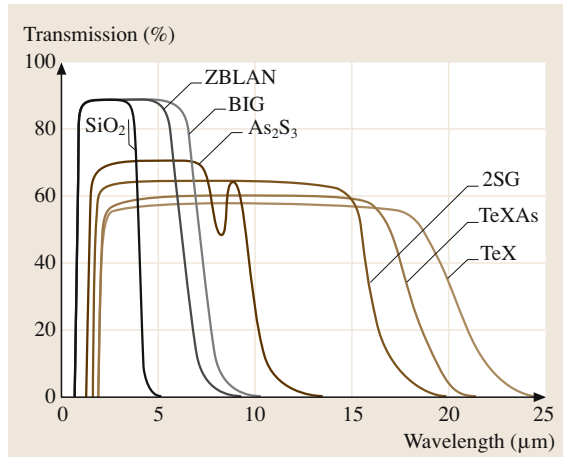


Fig. 43.1 Transmission plots are shown for SiO_2 glass and some representative fluoride and chalcogenide glasses. ZBLAN and BIG are heavy-metal fluoride glasses, 2SG is a selenide glass, and TeXAs and TeX are tellurium-based chalcogenide glasses. (After [43.7])

are globally associated with glassy materials. Further, an emphasis is placed on the unique advantages (lower processing temperatures, higher refractive index contrast, active functionality, etc.) enabled by some glass systems.

With these goals in mind, the chapter is organized as follows. In Sect. 43.1, we highlight the main advantages of glasses for photonic integration. These attributes include the unique processing options afforded by the existence of the glass transition and by the metastability of glasses. Section 43.2 provides an overview of the most-studied materials for glass-based integrated optics. This includes the standard glass formers (oxide, halide, chalcogenide glasses) as well as several materials that are generally classified as ceramics (Si_3N_4 , Al_2O_3 , etc.) but which can be realized as amorphous thin films. In Sect. 43.3, some unique advantages of glasses for realization of integrated light sources are reviewed. Section 43.4 provides a summary and some thoughts on future prospects for glasses in photonic integration.

43.1 Main Attributes of Glasses as Photonic Materials

The essential characteristic of a glass is the existence of a glass transition temperature (T_g). In short, if a viscous liquid is cooled fast enough, the liquid passes through the freezing temperature of the material with-

out making a phase transition to the solid crystalline state (see Fig. 43.2). Rather, it becomes a supercooled liquid, whose rate of change of specific volume (or enthalpy) with decreasing temperature remains constant

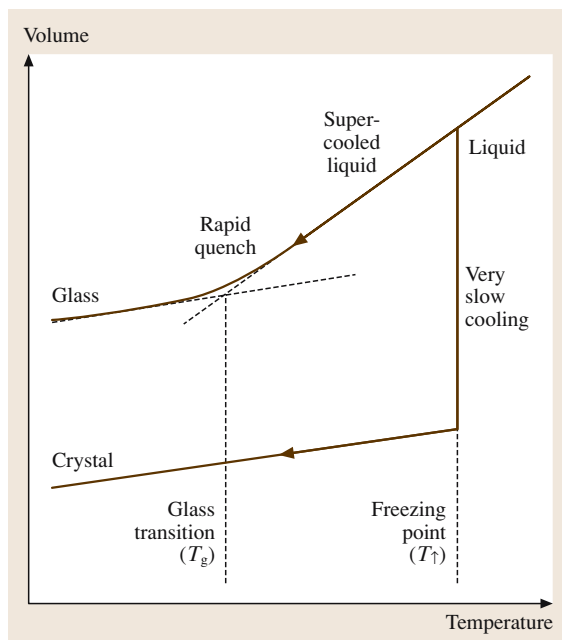


Fig. 43.2 Schematic plot illustrating the change in specific volume of a glass with temperature (after [43.11])

and equivalent to that of the molten liquid. At T_g , the slope of these curves (for example, specific volume versus temperature) changes abruptly (but in a continuous way) to a value that is approximately that of the corresponding crystal. Below T_g , the material is an amorphous solid or glass.

While glass forming has been a staple of human technological innovation for thousands of years, the physical processes underlying the glass transition are still a matter of some debate [43.11–13]. Several empirical facts are well known, however. The glass transition temperature of a particular material is not a constant, but is somewhat dependent on processing details such as the rate of cooling from the melt, or more generally on the technique used to form the glass (melt quenching, thin film deposition, etc.) and even the age of the glass [43.14]. T_g is slightly higher for a faster rate of cooling from a melt, for example, but in general $T_g \approx 2T_m/3$, where T_m is the melting/freezing temperature of the material [43.12]. For glasses that cannot be formed easily by melt quenching, such as amorphous Si [43.13], the direct experimental determination of T_g is not always possible, and indirect methods are sometimes used to estimate an effective T_g .

Central to the technological application of glasses is the ability to control their viscosity by way of temperature [43.5]. For stable glass formers, it is possible to

transition between states of very low viscosity (molten liquid), low viscosity (supercooled liquid), and high viscosity (frozen liquid; i.e. glass). This remarkable variation in mechanical properties with temperature, spanning several orders of magnitude in viscosity, underpins both the ancient and modern manufacture of glass products (by molding, extrusion, float processes, etc.). It is interesting to note that T_g can be defined as the approximate temperature at which the shear viscosity attains 10^{12} Pa s. Further, viscosity varies rapidly in the vicinity of T_g . For so-called strong glass formers (such as SiO_2), the viscosity η varies according to an Arrhenius equation of the form $\eta = A \exp(B/k_B T)$, where A and B are temperature-independent constants and k_B is the Boltzmann constant [43.12]. To form glasses into desired shapes, the temperature is typically controlled such that the viscosity is in the 10^2 to 10^7 Pas range, depending on the process and the stability of the particular glass composition [43.6, 15].

It was once held that only certain materials could form glasses, but it is now believed that the glassy state is attainable for all types of materials [43.11]. Accordingly, the literature contains references to various families of glasses, including inorganic (ceramic), organic (polymeric), and metallic glasses.

43.1.1 The Glass Transition as Enabler

The surface tension of liquids is a well-known enabler of optical quality surfaces [43.16]. Glass devices can be formed in the liquid state, and can thus accommodate manufacturing processes that are assisted by surface tension mediated self-assembly. The manufacture of modern single mode fibers is a good example of the enabling nature of the glass transition. In a typical process [43.4, 6, 17], highly purified glass layers are deposited either on the inside walls of a fused silica tube by chemical vapor deposition (CVD) or on the outside surface of a ceramic rod by flame hydrolysis deposition (FHD). The hollow glass cylinder (after removal of the ceramic rod in the latter case) is then heated (sintered) to $\approx 1600^\circ\text{C}$, such that it collapses to form a solid glass preform (typically 1 m in length and 2 cm in diameter). 1600°C is the approximate softening point of silica glass, typically defined as corresponding to a viscosity of $\approx 10^7$ Pas [43.6]. The fiber drawing process involves local heating of the preform to a temperature of approximately 2000°C (corresponding to a viscosity of $\approx 10^5$ Pa s). Thus, viscosity control is critical at nearly all stages of the process. Further, the drawing process allows tremendous control over the optical and

geometrical properties of the fiber, in part due to the self-regulating properties of surface tension effects in the supercooled liquid. As an added benefit, inhomogeneities in the original preform tend to be averaged out in the process [43.18]. The amazing properties of silica fibers (tensile strength comparable to that of steel, low loss, and extremely tight tolerances for material and geometrical parameters) are well known, and attest to the technological versatility of (silica) glasses. It is also interesting to note that the standard technique for splicing sections of silica fibers together, which is achieved with less than 0.1 dB loss [43.5], is based on a reflow process wherein surface tension contributes to their alignment.

Within integrated optics, the manufacture and optimization of PLC devices relies directly on the glass transition in many cases [43.9, 10]. PLC devices are manufactured in doped SiO₂ glasses deposited (by CVD, FHD, or sol-gel techniques) on a silicon wafer. The undercladding is often pure SiO₂, which can be formed by thermal oxidation of the silicon wafer. Typically, dopants in the core and uppercladding layers act to modify both the refractive index and the thermo-mechanical properties of the glass. In particular, controlled reflow (by heating the glasses near T_g) of these layers often plays an important part in the process. Reflow is used to consolidate the interface between layers, to smoothen the rough sidewalls of the waveguide core (formed by dry etching processes), and to assist in gap filling and planarization of the cladding layer over the patterned core layer [43.9], [43.18–20]. Typically, the layers are doped such that the glass transition decreases in the upward direction. In that way, a layer can be reflowed without affecting the underlying layers. As an example, a ‘pinned based reflow’ is shown in Fig. 43.3 [43.19]. The circle-section shaped core arises from selective reflow of the core material while its interface with the undercladding glass remains fixed. The upper boundary of the core is shaped by surface tension.

Surface tension driven manufacture of more exotic optical components has been widely studied. Examples include silica-glass based optical microcavities [43.21–23], nm-scale silica wires [43.24], photonic crystal fibers [43.25], microball lenses [43.26], and microlenses [43.27, 28]. In these cases, the reduction of scattering losses by surface tension mediated smoothening of interfaces is particularly important.

As discussed further in Sect. 43.2, high index contrast waveguides, photonic crystals, and microcavities are major themes within integrated optics. High index contrast (between the core and cladding of a waveguide, for example) is desirable because it enables strong con-

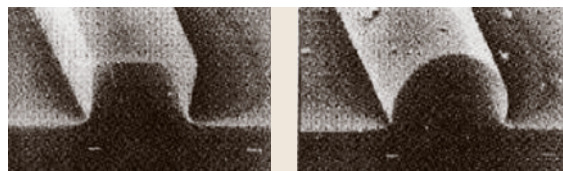


Fig. 43.3 A glass waveguide core is shown before (*left*) and after (*right*) a selective reflow process (From [43.19])

finement of light to small waveguides or small resonant cavities. This is essential to the realization of high-density integrated optics [43.9]. The downside of high index contrast is that scattering at interfaces increases with index contrast. This is clarified by the so-called Tien model for waveguide loss due to interfacial scattering in a symmetric slab waveguide [43.29]:

$$\alpha_S = \frac{2\sigma^2 k_0^2 h}{\beta} \Delta n^2 E_S^2 \quad (43.1)$$

where σ is the standard deviation (characteristic amplitude) of the roughness, k_0 is the free-space wavenumber, h is the inverse of the penetration depth of the mode into the cladding, β is the propagation constant of the mode, and E_S is the normalized electric field amplitude at the core-cladding interface. Further, $\Delta n^2 = n_{\text{core}}^2 - n_{\text{clad}}^2$, where n_{core} and n_{clad} are the core and cladding refractive indices, respectively. This equation explicitly shows the strong dependence of scattering loss on index contrast. A more accurate expression, derived by considering the statistics of roughness (not just the amplitude), is the Payne-Lacey model [43.30]:

$$\alpha_S = 4.34 \frac{\sigma^2}{k_0 \sqrt{2} d^4 n_{\text{core}}} \cdot g \cdot f \quad (43.2)$$

where d is the core thickness, g is a function that depends mainly on the core-cladding index offset, and f is a function that depends on the statistics of the roughness. If $r(z)$ is the random function describing the line-edge roughness (i.e. the deviation of the core-cladding interface from a straight line), the autocorrelation of $r(z)$ is often well described by an exponential function [43.31]:

$$R(u) = \langle r(z)r(z+u) \rangle \approx \sigma^2 \exp\left(\frac{-|u|}{L_c}\right) \quad (43.3)$$

where L_c is the so-called correlation length. It can be shown that for waveguide geometries of typical interest, low scattering loss requires both σ and L_c to be low. As a second illustration of the impact of interface roughness, it can be shown that the surface scattering limited quality (Q) factor for whispering gallery modes in a spherical

microsphere is approximately [43.23]:

$$Q_{ss} = \frac{\lambda^2 D}{2\pi^2 \sigma^2 L_c} \quad (43.4)$$

where D is the microsphere diameter. Again, minimization of scattering losses is correlated with minimization of σ and L_c .

It is interesting to compare recent results for high index contrast structures in crystalline materials and glasses. Much of the work on very highly confining waveguides and photonic crystals has focused on semiconductor material systems, especially the silicon-on-insulator (SOI) system [43.29, 30] and III-V semiconductor systems [43.33]. Generally, structures are defined by electron-beam or photolithography, followed by a dry etching process. Each of these steps has limited precision and contributes to the overall roughness [43.31]. With use of sophisticated processes, roughness parameters are typically on the order of $\sigma \approx 5$ nm and $L_c \approx 50$ nm, resulting in losses on the order of 3–10 dB/cm for sub-micron strip waveguides [43.30]. Significant improvement in these numbers is mostly reliant on advancements in lithography or etching processes. These same issues limit the performance of other high index contrast structures formed in semiconductors. For example the Q factor of semiconductor microdisk and microring resonators typically does not exceed 10^5 [43.34], and the loss of photonic crystal defect waveguides in semiconductors remains impractically large [43.35].

By comparison, glasses (at least certain glasses) offer the potential for much smaller characteristic roughness. If a glass is formed from a supercooled liquid state, surface tension effects can produce glass surfaces with nearly atomic level smoothness. The roughness of a melt-formed or reflowed glass surface is determined by surface capillary waves, which are small amplitude fluctuations at a liquid surface that become frozen in place at T_g [43.25]. The resulting RMS roughness is approximately given by [43.36]:

$$\sigma \approx \sqrt{\frac{k_B T_g}{\gamma(T_g)}} \quad (43.5)$$

where $\gamma(T_g)$ is the surface tension of the supercooled liquid at T_g . For silica glass, this equation predicts rms roughness on the order of 0.1 nm or less, and the correlation length of this inherent roughness is estimated to be as small as 3 nm [43.23]. Note that glasses with lower T_g and/or higher surface tension can in theory exhibit even lower values of surface roughness. Interestingly, sub-nm surface roughness has been experimentally verified

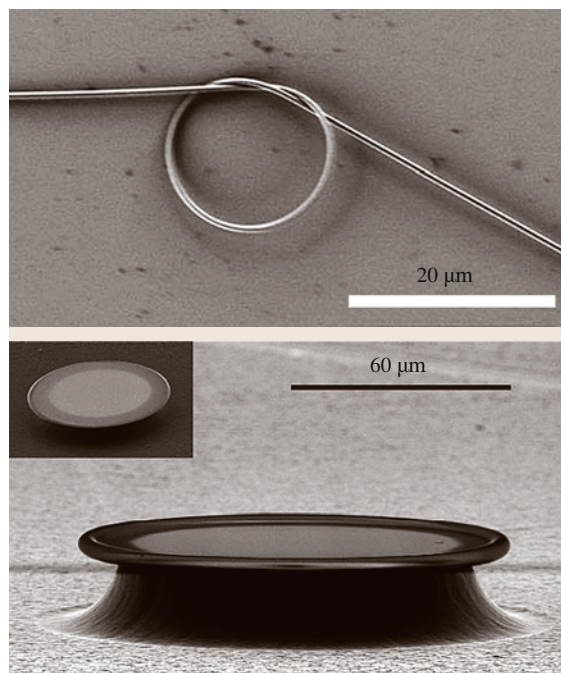


Fig. 43.4 (a) A submicron silica glass wire with sub-nm surface roughness is shown [43.24]. (b) A silica glass toroid sitting atop a silicon post is shown. The atomic-level smoothness of the toroid surface was achieved using a selective reflow process. (After [43.32])

for both melt-formed and fractured surfaces of silica glass [43.36] and indirectly from loss measurements in photonic crystal fibers [43.25]. Sub-nm roughness has also been determined for wet-etched surfaces of SiO_2 glass [43.37] and is not untypical for glass films deposited by evaporation.

The smoothness of glass surfaces has important implications for realization of low loss microphotonic devices, as evidenced by several recently reported results. While little work has been conducted on small core, high index contrast planar waveguides in glass, a method for fabricating sub-micron silica wires was recently reported [43.24] (see Fig. 43.4a). The authors estimated the rms surface roughness of their wires to be less than 0.5 nm, and demonstrated losses on the order of 1 dB/cm at 633 nm for 400 nm diameter wires. Sub-nm surface roughness has also been measured in the case of sol-gel glass microlenses formed by a reflow process [43.27]. Silica glass based microsphere cavities, typically formed by melting the end of fused silica fiber, exhibit Q factors of nearly 10^{10} [43.22, 23], the highest for any solid state microcavity. Recently, this concept

has been extended [43.21, 32] to the manufacture of silica toroid microresonators on a silicon substrate (see Fig. 43.4b). The latter devices exhibit Q factors exceeding 10^8 . Finally, glass-based photonic crystal fibers have been demonstrated to exhibit scattering loss near fundamental limits set by surface capillary waves [43.25]. This last result contrasts with the situation mentioned above for semiconductor photonic crystals, which to date are significantly compromised by optical scattering.

In summary, processing enabled by the glass transition clearly offers unique benefits for manufacture of microphotonic structures. It is necessary to add a caveat to this discussion, however. Many amorphous materials are conditional glass formers, and do not offer the full range of processing advantages outlined above. As an extreme example, amorphous silicon films typically crystallize at temperatures well below their effective glass transition temperature [43.13], and are not amenable to reflow processes. To varying degrees, this is true of many ‘glassy’ films used in integrated optics, such as silicon oxynitride and alumina. Further, the tendency towards crystallization varies greatly even amongst the traditional glass formers, as discussed below. Thus, chalcogenide and fluoride glasses will not generally support the range of viscosity control that enabled many of the SiO_2 -based devices mentioned. Having said that, novel techniques can sometimes circumvent crystallization problems. For example, pressure can be used as a tool for forming glass devices at temperatures below the onset of severe crystallization (but high enough to enable the glass to flow). Chalcogenide glass based lenses molded by the application of pressure and temperature are now commercially available [43.38]. Another interesting example is the recent work on extruded channel waveguides reported by Mairaj et al. [43.39]. Finally, much research has been directed at the fabrication of microphotonic elements in organic glasses by hot embossing [43.8]. In this technique, a hard master (such as a silicon wafer) is pre-patterned with the negative image of the desired photonic circuitry. The glass is heated above T_g and the master is pressed against the glass such that the image of its pattern is transferred to the softened glass. While studied mainly in polymer, this technique might also be applicable to low cost fabrication of photonic structures in ‘soft’ inorganic glasses.

43.1.2 Metastability

A glass is a metastable material, having been frozen as an amorphous solid possessing excess internal en-

ergy relative to the corresponding crystal. Metastability is a double-edged sword. On the one hand, it presents technological limitations with respect to the processing and use of a glass device. As an example, it is well known that commercial introduction of organic glass (polymer) based optical devices has been hindered by poor stability of many polymers at elevated temperatures or under exposure to high intensity light. On the other hand, metastability offers unique options for fabrication and optimization of photonic microstructures. Specifically, glass properties (optical, mechanical, chemical) can often be tailored or adjusted by the careful addition of energy, in the form of heat, light, electron beams, or ion beams. An example is provided by the commercially important fiber Bragg gratings (FBG), wherein UV light is used to induce a stable refractive index change in the Ge-doped core of a standard SiO_2 -based fiber. In the following subsections, the implications of metastability for glass-based integrated photonics are discussed.

Devitrification

Well below its glass transition temperature, a glass sits in a local energy minimum and is impeded by its own viscosity from reaching the lower energy crystalline state. A frequently cited example is that of silicate glass windows in ancient buildings, which have not exhibited any significant crystallization. However, addition of sufficient energy (such as heating a glass to some temperature above T_g) can result in crystallization of a glass. Controlled crystallization is a technologically useful means of modifying the mechanical, thermal, or optical properties of a glass, used in the production of glass-ceramics. However, devitrification produces a polycrystalline material, increases optical scattering loss, and is usually a problem to be avoided in manufacture of photonic devices.

As mentioned in Sect. 43.1.1, crystallization tendencies vary widely between various glass systems. The difference between the onset temperature for crystallization and T_g , $\Delta T = T_x - T_g$, is one of the parameters that define the stability of a particular glass. The high stability against devitrification of silicate glasses is one of numerous reasons that they have traditionally dominated glass technology. This stability underlies the manufacture of the structures discussed in Sect. 43.1.1. Other glass forming systems have lower stability against devitrification, often making their technological application a greater challenge [43.39, 40]. For example, a major research thrust in the field of fluoride and chalcogenide glass fibers is the identification of new compositions that

exhibit improved stability against crystallization during the fiber drawing process [43.41].

For integrated optics, this form of stability is not always so critical. Amorphous thin films created by chemical vapor deposition, sputtering, evaporation, etc. can often be processed at temperatures below the onset of any significant crystallization. This enables the use of conditional glass formers, such as Si_3N_4 , Al_2O_3 and semiconductors, in their amorphous state. However, it is interesting to review the early development of integrated optics, much of it focused on the identification of suitable thin film materials. Studies [43.43–45] that compared waveguide fabrication in polycrystalline (ZnO , ZnS , Al_2O_3 , etc.) and amorphous (Ta_2O_5 , polymer, sputtered glass) thin films showed that losses were orders of magnitude higher for the polycrystalline films. The lowest losses were initially obtained for polymer films [43.44], attributed to their amorphous structure and the smoothness of surfaces arising from spin casting of films.

If present, uncontrolled crystallization within films can be a dominant source of propagation loss [43.46]. For conditional glass formers, such as Al_2O_3 and TiO_2 , crystallite formation must be carefully avoided during the deposition process [43.47]. While less of an issue for natural glass formers, the possibility for crystallization must always be considered when a glass film is subjected to high temperature processing steps.

Structural Relaxation

Having been frozen in some metastable configuration at the time of formation, an amorphous solid will always be subject to some degree of aging (changes in the material properties on some time scale). At temperatures below T_g , the glass is usually inhibited from crystallization by a significant internal energy barrier. However, subtler changes in the network structure can occur, such as the transition from the initial state to a more stable (but still amorphous) second state. This so-called structural relaxation is manifested in many ways – experiments often probe the change in time of specific volume, structural signatures (such as by Raman spectroscopy), or optical properties of the glass. The rate of structural relaxation is highly dependent on temperature, being relatively fast at temperatures near T_g and significantly slower for temperatures far below T_g [43.13]. In fact, annealing of glasses at a temperature near T_g is a standard technique used to promote rapid structural relaxation and lessen the dependence of material properties on the processing history [43.6]. In many cases, this annealing step is not practical, and structural relaxation over the in-use lifetime of a glass-based device must be considered.

As the following discussion will illustrate, the rate of structural relaxation at temperature T_1 exhibits a logarithmic dependence on the temperature difference $T_g - T_1$. Thus, T_g is an important parameter in character-

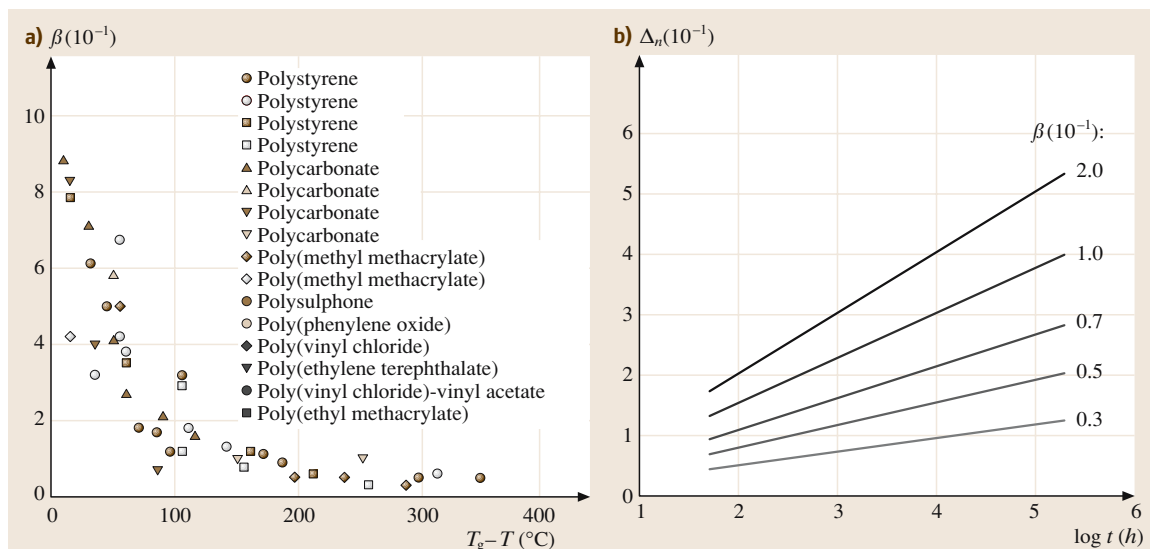


Fig. 43.5 (a) A plot of the structural relaxation rate parameter versus the supercooling temperature for a wide range of organic glasses is shown. (b) The predicted change in refractive index versus time for an organic glass, with the rate constant β as a parameter. (After [43.42])

izing the long-term stability of a glass component. For example, SiO_2 has a very high T_g ($\approx 1100^\circ\text{C}$) and does not exhibit significant relaxation at room temperature, either on the time scale of typical experiments or over the typical lifetimes envisioned for glass devices. In fact, aging effects in silica fibers are predominately associated with the growth of defects formed in the glass at the time of manufacture. Environmental moisture is a particular contributor to this aging process. On the other extreme, structural relaxation in low T_g organic glasses is known to be an issue, and intrinsic aging effects in polymers have been widely studied [43.48]. In particular, low T_g glasses exhibit densification with time (relaxation of their specific volume) and a corresponding change in their refractive index. This is a particular concern for interferometric optical devices (gratings, resonators, Mach–Zehnder interferometers), which necessitate tight control of material indices.

Volume relaxation can be characterized by a rate parameter β , defined as:

$$\beta = \frac{1}{V} \left[\frac{\partial V}{\partial (\log t)} \right]_{P,T} \quad (43.6)$$

where V is volume, t is time, P is pressure, and T is temperature. Using the Lorentz–Lorenz expression, the following relationship has been derived from (43.6) [43.42]:

$$\left[\frac{\partial n}{\partial (\log t)} \right]_{P,T} = \frac{-\beta \left(\frac{\partial n}{\partial T} \right)_{P,T}}{\alpha} \quad (43.7)$$

where n is refractive index, $\partial n / \partial T$ is the thermo-optic coefficient, and α is the volume coefficient of thermal expansion. Zhang et al. [43.42] further showed that, for polymers, β is approximately a universal function of $T_g - T_1$ as shown in Fig. 43.5. This implies that the time rate of change of refractive index due to structural relaxation is strongly correlated with the parameter $\Delta T_{SC} = T_g - T_1$, termed the supercooling temperature. They also assessed the implications for polymer-based telecommunications devices, concluding that T_g needs to be higher than 300°C for many applications. This result provides a nice illustration of the impact of T_g on device stability.

While it is not necessarily reasonable to extend the foregoing results directly to inorganic glasses, the essential features of structural relaxation will be similar. The chalcogenide glasses in particular are sometimes called inorganic polymers, partly because of their softness and mechanical flexibility relative to silica-based glasses. Further, the range of T_g for chalcogenide glasses

(≈ 50 – 550°C) is similar to that of organic polymers, and studies on chalcogenide glasses [43.14] support the conclusion that structural relaxation would present similar restrictions on inorganic and organic glasses. Thus, lifetime restrictions due to relaxation should be considered in the application of any glass having relatively low T_g , including chalcogenide glasses, fluoride glasses, and many non-silicate oxide glasses [43.6].

Both the processing temperatures (during manufacture) and the in-use temperatures are important considerations in assessing the stability of a glass. Telecommunication devices are typically designed to withstand in-use temperatures between -40°C and 85°C [43.49]. As an example of a more demanding application, it has been proposed that glasses are good candidates for fabrication of integrated photonic elements on future integrated circuit chips [43.50]. In-use temperatures on modern microprocessors can reach 100 – 200°C .

Finally, it should be noted that any energy applied to the glass (light, electron beams, etc.) might induce structural relaxation. This has been most widely studied for the chalcogenide glasses, which exhibit a wide array of photoinduced structural changes [43.51]. A fascinating example is the so-called photoinduced fluidity phenomenon [43.52], in which intense sub-bandgap light can reduce the viscosity of a chalcogenide glass by several orders of magnitude (causing the glass to flow or melt). This effect is completely athermal, and in fact is often enhanced at low temperatures. While an extreme example, it illustrates the possibility of aging effects due to non-thermal processes in glass devices. For photonic devices, the main concern is usually the effect of light exposure over the in-use lifetime.

Metastability as Enabler

One highly desirable implication of metastability is the possibility for direct patterning of photonic structures in glasses using energetic beams (electron beams, ion beams, light). These effects can also be exploited for post-fabrication trimming of devices [43.53], which is important since many optical components have tolerances beyond the capabilities of practical fabrication processes.

Beam induced effects in glasses are generally linked to their random network structure. That is, unlike crystals, glasses contain a range of internal ‘defects’ – wrong bonds, dangling (missing) bonds, impurity atoms that act as network modifiers, as well as their inherent variation in bond angles. Addition of energy can cause a glass to undergo a transition from one metastable state to another. This transition is often accompanied by a change in the

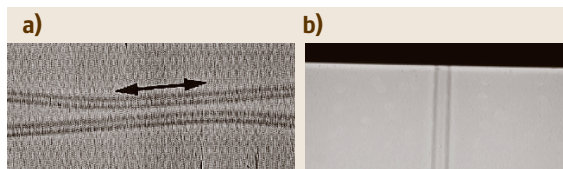


Fig. 43.6 (a) A directional coupler is shown, as written in a multicomponent silicate glass by an ultrafast Ti:sapphire laser [43.54]. The black arrow indicates the coupling region. (b) Microscope image of a rib waveguide formed by direct exposure of a chalcogenide glass (As₂Se₃) in a UV mask aligner, followed by wet etching [43.55]

density of defects or by a structural change – rearrangement or reorientation of bonds. Associated with these structural changes are changes in the physical, chemical, and optical properties of the glass. Most existing knowledge about these induced structural changes is of an empirical nature, whereas the underlying physical mechanisms are still the subject of research.

The ability to induce transitions between metastable states is probably a universal property of amorphous matter. Perhaps for practical reasons, such as the availability of light sources of suitable wavelength, metastability was initially studied in amorphous semiconductors and low band gap chalcogenide glasses. It should be added, though, that chalcogenide glasses seem to be unique with respect to the variety of induced structural changes they exhibit or at least the magnitude of those changes [43.51]. This might be related to the relatively weak, highly polarizable covalent bonds that are characteristic of chalcogenide glasses. Nevertheless, the later discovery [43.56] of photoinduced refractive index changes in Ge-doped silica glass prompted a huge research and development effort because of the technological importance of silicate glasses. The main outcome of this work was the fiber Bragg grating (FBG), which is a critical component in modern telecommunications and sensing networks. FBGs are manufactured by inducing a quasi-permanent index change within the Ge-doped core of a standard silica fiber, by exposure to an UV laser beam. Writing wavelengths are in the 150 to 350 nm range, corresponding to photon energies below the nominal bandgap of SiO₂. The mechanisms underlying the index increase are still debated [43.9, 56]. However, it is believed that contributing factors are absorption by defects in the glass

(possibly generating further defects) and densification due to some sort of structural relaxation [43.57]. Typical index changes for standard fiber cores are in the 10⁻⁴ to 10⁻³ range, but several techniques (addition of other dopants, hydrogen-loading, etc.) have been developed to enable changes as high as 10⁻². Given the discussion in the preceding section, it is reasonable to ask whether photoinduced changes are themselves stable. The answer, in some cases, seems to be yes. Researchers have concluded that fiber Bragg gratings can have an operational lifetime of greater than 20 years, even at elevated temperatures of 80 °C [43.58].

Photosensitivity has been widely explored as a means to directly pattern integrated optics devices, especially in silica-based glasses [43.59]. In general, direct patterning is induced using photon energies close to the nominal bandgap energy of the glass or using intense light of lower photon energy [43.57]. In the latter case, photoinduced effects arise from nonlinear effects such as multi-photon absorption. In particular, the widely available Ti:sapphire laser has been applied to the writing of microstructures in a variety of glasses [43.54]. Because of the nonlinear nature of the writing process, writing with sub-band gap light allows the direct patterning of 3-dimensional structures (see Fig. 43.6).

Photosensitivity has been demonstrated within all of the standard glass families, and in many other amorphous thin films (see Sect. 43.2). Because of their large photosensitivity, direct patterning of integrated optics structures in chalcogenide glasses is possible using low power light sources [43.55, 60].

43.1.3 Glass as Host Material

For completeness, it should be mentioned that glasses are unique in their ability to incorporate a wide range of dopants, sometimes in very high concentration. This is related to the random network structure of a glass. Foreign species are much more likely to find a suitable bonding site, or simply space to reside, inside a glass versus a crystal. These foreign species can be rare-earth ions, transition metal ions, or semiconductor or metal nanoparticles, for example. As a result, a glass provides significant scope for active functionality. We can transform a passive glass into a laser glass, a nonlinear optical glass, or a magneto-optic glass, for example, by appropriate doping [43.3, 6].

43.2 Glasses for Integrated Optics

In general, thin films can be amorphous, polycrystalline, or crystalline. *Tien* and *Ballman* [43.45] provided an early review of waveguide results achieved for various materials lying within each of these categories. Unless the crystal grain size is sufficiently small relative to the wavelength of interest [43.47], polycrystalline films are too lossy for integrated optics. It should be noted that high attenuation might be tolerable if circuit length is sufficiently short. The present discussion is concerned with amorphous films, which *Tien* further subcategorized as low index ($n < 1.7$), medium index ($n < 2$), and high index ($n > 2$). For convenience, we will follow a similar approach in the following sections. This categorization is somewhat arbitrary, especially when considering material systems (such as the silicon oxynitride system) that enable a range of refractive index. It is interesting to note that 1.7 is the approximate upper limit for the refractive index of organic glasses [43.49]

As noted in Sect. 43.1, increased circuit density is one of the primary goals of integrated optics research [43.9]. Whether employing traditional total internal reflection effects or photonic band gap materials, increased density relies on high index contrast between at least two compatible materials. Note that the terminology ‘high index’ and ‘high index contrast’ are rather imprecise. For example, in silica PLC technologies core-cladding index differences of $\Delta n \approx 0.02$ have been labeled as ‘superhigh’ index contrast [43.10]. This is a very small value, however, relative to the index contrasts that characterize SOI photonic wire waveguides [43.30].

43.2.1 Low Index Glassy Films

SiO₂-Based Glasses

As mentioned in the introduction, the SiO₂ on silicon PLC system is the most widely used and developed glass-based integrated optics platform. PLC development (mainly in the 1990s) was driven by fiber optic long haul communication systems, especially the emergence of wavelength division multiplexing (WDM). Various devices, but especially arrayed waveguide grating (AWG) wavelength demultiplexers, were developed to a very high degree of sophistication by the telecommunications industry. Good early [43.18] and recent [43.9,10] reviews are available in the literature. A brief overview of the technology is given below, as it illustrates some of the advantages and challenges associated with glass-based photonic integration.

As mentioned in Sect. 43.1.1, commercial PLC waveguides are fabricated primarily by FHD [43.10] or by CVD [43.9]. In addition, considerable research has been conducted on sol-gel synthesis [43.61], with the aim of reducing fabrication costs and providing greater flexibility over the choice of glass compositions. Typically, the undercladding is pure silica glass. To raise its refractive index, the core layer is doped with Ge or P. Ge-doped SiO₂ is known for its photosensitivity, as discussed in Sect. 43.1.2. Addition of P lowers the viscosity and characteristic reflow temperature, enabling the processing options discussed in Sect. 43.1.1. Typical relative index offset Δ between the core and cladding is in the 0.3 to 2% range, corresponding to minimum waveguide bend radius R in the 2 to 25 mm range [43.10]. After RIE to form an approximately square waveguide core, an upper cladding is deposited and subjected to heat treatment. The upper cladding is often a boro-phosphosilicate glass (BPSG), partly to enable reflow. Further, the B and P dopants have opposite effects on refractive index allowing a nearly symmetric waveguide structure to be obtained.

Given the target applications, it is not surprising that PLC technology placed a great emphasis on efficient coupling between the integrated waveguides and external fiber waveguides. Thus, waveguide cross-sectional dimensions and refractive index contrast between core and cladding layers were tailored to provide a good impedance match (low reflection and good modal overlap) to standard fiber. By employing essentially the same glass (doped SiO₂) as that used to construct fiber, it is even possible to achieve intimate reflowed (fusion type) coupling between the integrated and fiber guides [43.18]. The emphasis on impedance matching is due to the great importance of minimizing insertion loss in fiber systems, but presents some practical limitations:

1. The relatively low index contrast ($\approx 10^{-3}$) between core and cladding necessitates thick glass films. For example the undercladding or buffer layer, often a thermally grown SiO₂ layer, must typically exceed 12 μm in order to negate radiation losses into the high index silicon substrate. Further, high temperature anneals (typically 900–1150 °C) are required to drive out hydrogen impurities and to reflow the core and upper cladding layers. The combination of thick films, high temperature anneals, and thermal expansion mismatch can result in wafer bending and damage to the glass films. This is partly alleviated

in practice by depositing identical layers on both sides of the Si wafer [43.9], which adds cost and complexity.

2. Since modal area and minimum bend radius scale inversely with core-cladding index contrast, traditional PLC waveguides do not support high-density optical integration. As mentioned above, index contrasts in the 0.3 to 2% range correspond approximately to bending radii in the 25 to 2 mm range [43.10].

In recent work at Corning [43.62], very low loss (< 0.1 dB/cm) waveguides and ring resonators were realized in PECVD grown silica-germania waveguides having index contrast as high as 4%. Such high index contrasts can accommodate bending radius of less than 1 mm, which is comparable to the range explored recently by IBM researchers using silicon oxynitride materials [43.63].

Amorphous Aluminium Oxide

Sapphire (crystalline Al_2O_3) is amongst the most important solid-state laser hosts. Alumina (polycrystalline or amorphous Al_2O_3) is an important industrial material in its own right, possessing outstanding mechanical and thermal properties [43.15]. Amorphous Al_2O_3 films are of interest for integrated optics for several reasons [43.47]. First, the refractive index is relatively high (although falling within the low index range specified above), typically $n \approx 1.65$. Second, Al_2O_3 is an excellent host for rare earth and transition metal dopants. As discussed in Sect. 43.3, rare earth dopants of interest are typically trivalent, matching the valency of Al ions in Al_2O_3 [43.64]. As a result, rare earth ions can be incorporated easily into the alumina matrix. In short, Al_2O_3 can homogeneously dissolve large concentrations of rare earth ions and is therefore of interest for realization of integrated amplifiers and light sources. Finally, Al_2O_3 films typically have excellent transparency from the UV to mid-IR range.

Fluoride Glasses

Heavy-metal fluoride glasses (typically fluorozirconate glasses) such as ZBLAN have been widely studied since their discovery in 1974 [43.41]. One of their outstanding attributes is a wide transparency range, extending from the UV well into the mid-IR (see Fig. 43.1). They have received considerable attention as fiber optic materials, because theory predicts a minimum absorption well below that of silica glass. Further, rare earth ions exhibit the greatest number of useful radiative transitions when embedded in fluoride glasses [43.65]. This is due to

their wide transparency window and low characteristic phonon energies (see Sect. 43.3).

Fluoride glasses typically have refractive indices in the 1.47–1.57 range, which is advantageous in terms of being well matched to silica glasses. Thin film deposition of these complex multicomponent glasses is difficult, and thermal expansion mismatches with standard substrates create further challenges. For these reasons, fluoride glasses have not been widely explored for applications in integrated optics. The so-called PZG fluoride glasses ($\text{PbF}_2 - \text{ZnF}_2 - \text{GaF}_3$) have been successfully deposited using straightforward evaporation techniques, enabling relatively low loss waveguides and erbium-doped amplifiers on a silicon platform [43.66].

43.2.2 Medium Index Glassy Films

Silicon Oxynitride

Silicon oxynitride (SiON) films are generally deposited by a CVD technique. The promise of this material system for integrated optics was identified in early work [43.45]. One of the main attributes of SiON is that it is a standard material system employed in silicon microelectronics, and the thin film technology has been developed accordingly [43.47, 63]. Because of this, SiON is currently being studied as potential material for on-chip interconnects [43.50]. Further, the system enables a continuous range of refractive index from approximately 1.45 (SiO_2) to 2 (Si_3N_4) at 1550 nm wavelength. This index range has been extended to approximately 2.2 by deposition of non-stoichiometric, silicon-rich nitride films [43.67]. SiON and SiN have been amongst the most explored materials for realization of microring resonator structures [43.53, 67].

As an optical material, SiON has some drawbacks. The main one is the presence of hydrogen impurities in films deposited by traditional techniques. Overtones due to hydrogen bonds (mainly N–H and O–H) can produce impractically large values of loss in the 1300 nm and 1550 nm telecommunications bands [43.62]. Long term, high temperature annealing (typically at $> 1000^\circ\text{C}$) is required to reduce this loss. Interestingly, structural relaxation on the time scale of hours and days has been observed in such annealed films [43.47]. Perhaps related to this metastability, UV light has been used as a means to trim the refractive index of SiN-based microring resonators [43.53]. Modified deposition processes that can produce SiON films having low stress and low hydrogen content (without requiring a high temperature annealing step) have been reported recently [43.63, 67]. However,

inherent film stress can limit the maximum thickness to a few hundred nanometers in some cases [43.50].

43.2.3 High Index Glassy Films

Amorphous Metaloxides

Many amorphous metaloxides (Y_2O_3 , Nb_2O_5 , Ta_2O_5 , etc.) have traditionally been used as high index layers in optical thin film stacks [43.49]. It is logical to look at these materials in the hunt for high index thin films for integrated optics. Often these oxides are miscible with SiO_2 , making it possible to tune the refractive index and other material properties over a range of values (as in the case of silicon oxynitride) [43.69].

TiO_2 films are used as dielectric layers, optical coatings, and to protect underlying materials from mechanical damage or corrosion [43.47]. The high dielectric constant of TiO_2 is widely exploited in microelectronics applications. It also has useful optical properties, including a refractive index as high as 2.6 (depending on film deposition details) and good transparency from the UV to mid-IR. The TiO_2 – SiO_2 material system has good properties for sol-gel synthesis of thin films [43.61]. Integrated waveguides based on TiO_2 have been studied [43.47, 69], and AWG devices in TiO_2 -rich oxide glass with $n \approx 1.9$ were recently reported [43.70].

Ta_2O_5 (With $n \approx 2.1$) is another well-studied material, used for example as a high dielectric constant material in microelectronics. It has low absorption in the visible to near infrared wavelength range, and has been recently studied as a material for active integrated optics [43.71, 72]. Further, $\text{Ta}_2\text{O}_5/\text{SiO}_2$ has been applied to the study of compact microring resonators [43.73].

TiO_2 and Ta_2O_5 have been used in the fabrication of three-dimensional photonic crystals with fundamental bandgaps in the visible and near infrared [43.68, 74]. An example is shown in Fig. 43.7.

Heavy Metal Oxide Glasses

These glasses are based primarily on the oxides of bismuth (Bi_2O_3) and lead (PbO). They have relatively low phonon energies and transmit further into the infrared than most oxide glasses [43.75]. They also have interesting magneto-optic [43.6] and nonlinear optical [43.3] properties.

Chalcogenide Glasses

Many chalcogenide alloys, based on S, Se, or Te, are excellent glass formers. In fact, Se is the only element able to easily form a glass on its own [43.14]. These

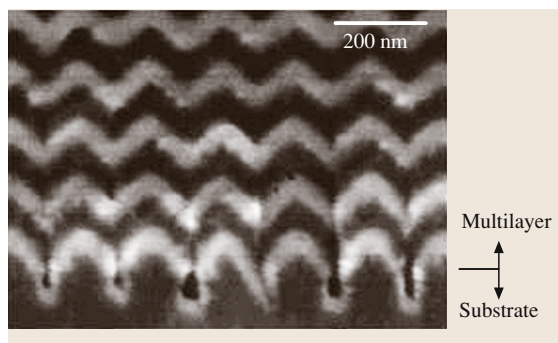


Fig. 43.7 A three-dimensional photonic crystal fabricated using the autocloning growth technique is shown. The *bright* and *dark* layers are TiO_2 and SiO_2 glass, respectively [43.68]

glasses are characterized by very low phonon energies, good transparency in the mid-infrared, and a wide array of beam-induced structural changes [43.51]. Further, chalcogenide glasses are essentially amorphous semiconductors, with electronic absorption edge typically in the visible to near infrared. Related to this, they have very high refractive indices, ranging from 2.2 for sulfide glasses to greater than 3 for some telluride glasses. They are currently of industrial interest for night-vision optics [43.38], fibers for infrared transmission [43.41], and optical and electronic memory elements [43.51].

Chalcogenide glasses have many appealing features for integrated optics. High quality thin films can often be deposited by straightforward techniques such as evaporation or sputtering. Their high indices make them suitable for realization of compact waveguides [43.55] and photonic crystals [43.76]. Their beam-induced properties enable many unique processing options. Some chalcogenide alloys have good potential as laser glasses [43.41], with the ability to uniformly dissolve large concentrations of rare-earth ions. Perhaps of greatest interest, the chalcogenide glasses are amongst the most promising materials for nonlinear integrated optics [43.77, 78]. On this last point, several chalcogenide glasses have ultrafast optical Kerr response 2 to 3 orders of magnitude higher than that of SiO_2 , while satisfying basic figure of merit criteria for all-optical switching. Results for metal-doped chalcogenide glasses [43.78] suggest even higher nonlinearities might be attainable. Combined with the possibility of realizing highly confining waveguides, this makes chalcogenide glasses one of few material systems that can realistically satisfy long-standing goals for nonlinear processing in integrated optical circuits. For example, it has been estimated [43.77]

that 1 pJ pulses of 1 ps duration (≈ 1 W peak power) could induce phase shifts (due to self phase modulation) greater than π in feasible chalcogenide glass waveguides with length on the order of a few centimeters. This is predicated on realization of low loss waveguides with modal area of $\approx 1 \mu\text{m}^2$. Recent experimental results suggest that this goal is within reach [43.55].

Tellurite Glasses

While TeO_2 is a conditional glass former, the addition of other oxides can result in stable glasses with

many interesting properties [43.75]. They typically have refractive index greater than 2, low phonon energy, large optical nonlinearities, and a high acousto-optic figure of merit. Further, they have been widely studied as hosts for rare-earth ions. Of particular interest has been the wide bandwidth of the 1550 nm emission exhibited by erbium in tellurite glass [43.79]. Because of these numerous attractive properties, tellurite glasses and amorphous TeO_2 films ($n \approx 2.2$) have received some attention for applications in integrated optics [43.80].

43.3 Laser Glasses for Integrated Light Sources

Active functionality includes means to generate and detect light (especially stimulated emission and absorption) and means to control (switch, modulate, etc.) light signals. Numerous material properties are employed in active photonics, including thermo-optic, acousto-optic, magneto-optic, electro-optic, and nonlinear Kerr effects. Many glasses have attractive thermo-optic or acousto-optic properties, and a few examples were cited in Sect. 43.2. Also in Sect. 43.2, some glasses with promising nonlinear optical properties were discussed. Elsewhere in this volume, K. Tanaka has provided an excellent review of nonlinearities in photonic glasses. A recent, thorough review of magneto-optic glasses is also available [43.6].

Arguably, the most critical element for photonic integration is an integrated light source. Rare-earth doped glasses are well-established laser media, used especially for the realization of bulk and fiber lasers. Considerable effort has been directed towards development of integrated amplifiers and lasers based on such glasses. In the following, we attempt to highlight some ways in which light sources based on glasses are uniquely enabling, relative to those based on crystalline materials. The performance advantages discussed below combined with the properties discussed in earlier sections (fabrication options) make glasses particularly attractive.

43.3.1 Advantages of Glass-based Light Sources

Stimulated emission devices in glass are almost always based on trivalent rare-earth dopants [43.6]. Thus, the term laser glass can usually be equated with the term rare-earth doped glass. Rare-earth doped laser glasses have been widely studied and reviewed [43.64, 65]. Fur-

ther, recent and comprehensive reviews on Nd- and Er-doped integrated glass amplifiers and lasers are available [43.6, 81]. Relative to single crystal hosts, glass hosts result in rare-earth ions exhibiting broadened luminescence lines and lower peak stimulated emission cross-sections [43.65]. This property is a result of the random network structure of glasses; embedded rare earth ions exist in a range of local environments. The broadened, weaker emission is of great importance for the realization of broadband, low noise fiber amplifiers.

It should be noted that the semiconductor injection laser is an extremely advanced technology, and is the dominant type of integrated light source at present and for the foreseeable future. Semiconductor lasers have important advantages over any glass-based device demonstrated to date. First, semiconductor gain media typically have gain coefficients of the order $\approx 100 \text{ cm}^{-1}$ [43.17]. By comparison, glasses require high rare-earth dopant concentration to achieve gain coefficients exceeding 1 cm^{-1} . Thus, semiconductor optical amplifiers (SOAs) and lasers have cavity lengths measured in tens to hundreds of μm while it is typical for glass waveguide amplifiers and lasers to be measured in cm. Second, semiconductor light sources are pumped electrically while glass devices are typically pumped optically. Electrical pumping is highly desirable for optoelectronic integration of photonic devices on electronic chips. However, as discussed below, integrated glass waveguide lasers have important advantages of their own [43.82].

The lifetime of the metastable lasing level in rare-earth doped glasses is usually on the order of ms, much longer than the ns lifetimes typical of semiconductor gain media. This long lifetime implies that the gain does not change rapidly with variations in input power (pump

or signal). This is an essential feature of the commercially important erbium-doped fiber amplifier (EDFA); the long lifetime (≈ 10 ms) of the $^4\text{I}_{13/2}$ level of erbium in silicate glass contributes to low noise operation, high pump efficiency, and low crosstalk between wavelength channels in a WDM system [43.83]. Further, since the relaxation oscillations in glass waveguide lasers occur at relatively low frequency, glass lasers can be modelocked at correspondingly much lower repetition rates compared to semiconductor lasers [43.82]. This can enable much higher peak intensities from the glass laser.

Related to the discussion in Sect. 43.1, lower cavity loss (higher cavity Q) is generally possible for glass devices. Further, the long metastable lifetime of the rare-earth transition allows glass lasers to have linewidths approaching the Schawlow–Townes limit [43.84]:

$$\Delta\nu = \frac{2\pi h\nu_0 (\Delta\nu_c)^2}{P} \left(\frac{N_2}{N_2 - N_1} \right) \quad (43.8)$$

where ν_0 is the laser center frequency, P is the laser output power, N_2 and N_1 are the population densities of the upper and lower lasing levels, and $\Delta\nu_c = (1/2\pi t_c)$ with t_c the photon lifetime in the cavity. The high Q of glass laser cavities coupled with relatively high output powers enables linewidths less than 10 kHz, orders of magnitude below that of semiconductor DFB lasers [43.82]. The high cavity Q and long metastable lifetime is also advantageous for achieving ultrastable passive mode locking, with low timing jitter and pulse-to-pulse power variation.

Finally, glasses offer the possibility of integration on various substrates. While semiconductor lasers are inherently integrated structures, they are not easily transportable between platforms. For example, III-V semiconductor lasers have not shown great promise (in spite of heroic efforts in some cases) to satisfy the desire for a compact, truly integrated light source on the silicon electronics platform. Glasses (perhaps doped with semiconductor nanocrystals) are increasingly viewed as the more promising route to achieving such a goal [43.64].

43.3.2 Alternative Glass Hosts

The theoretical maximum gain (cm^{-1}) of a rare-earth doped glass waveguide amplifier can be expressed [43.81] as $\gamma_p = \Gamma\sigma_p N_{\text{RE}}$, where σ_p is the peak (versus wavelength) stimulated emission cross-section (cm^2) for the transition of interest and N_{RE} is the volume density (cm^{-3}) of the rare-earth ions. Γ is a dimensionless factor (lying between 0 and 1) that accounts for the spatial overlap of the waveguide mode (at the wave-

length to be amplified) and the active ions producing the gain. It can be optimized through waveguide design, irrespective of the glass host, and will not be considered further here. The expression for γ_p neglects all waveguide losses (due to scattering, etc.) and assumes that all of the rare-earth ions have been promoted to the desired lasing level; i. e. a complete population inversion. It is therefore an ideal and unattainable limit, but is useful for framing the following discussion.

Since compactness is a central goal of integrated waveguide lasers, alternative glass hosts can be compared on the basis of the maximum gain (γ_{max}) that they enable in practice (for a given transition of a given rare-earth ion). Further, since low noise operation relies on a near complete population inversion [43.81], it is desirable that $\gamma_{\text{max}} \approx \gamma_p$. In simple terms, the glass should dissolve a large concentration of the rare earth ion (high N_{RE}), should result in a large stimulated emission cross-section for the desired transition, and should enable the realization of a nearly complete population inversion. The importance of other practical considerations, such as stability, processing options, and physical properties of the glass, will depend on the intended application. Some hosts provide unique advantages, such as flexible pumping options, as discussed in Sect. 43.3.3.

As mentioned, laser transitions in glasses are generally provided by radiative decay between two energy levels of a trivalent rare-earth ion. Perhaps the most important example is the $^4\text{I}_{13/2}$ to $^4\text{I}_{15/2}$ transition of Er^{3+} , which produces luminescence in the 1500–1600 nm wavelength range. Once an ion has been promoted (by pumping) to the upper lasing level, it will eventually transition to another state by interactions with the glass, impurities in the glass, the photon fields (at the signal or pump wavelength), or with other rare-earth ions in its vicinity [43.65]. The metastable lifetime of the upper lasing level can be expressed [43.81]

$$\frac{1}{\tau} = A + W_{\text{MP}} + W_{\text{ET}} + W_{\text{IMP}} \quad (43.9)$$

where A is the effective rate of spontaneous radiative decay to all lower lying levels, W_{MP} is the rate of non-radiative decay due to multi-phonon energy exchanges with the glass, W_{ET} is the rate of non-radiative energy transfer due to interactions between closely spaced rare-earth ions, and W_{IMP} is the rate of energy transfer to quenching impurity centers in the glass. The first three terms on the right will be discussed below. For the last term, a classic example is the quenching of the 1550 nm luminescence band of erbium due to resonant energy transfer to OH^- impurities in silica glass [43.65].

The choice of a particular glass host will impact several important properties of a given transition: pumping efficiency, peak gain, linewidth, metastable lifetime, etc. Alternative glasses can be compared on the basis of a few key parameters, as discussed in the following sub-sections. Representative data for erbium in various glasses is given in Table 43.1.

Stimulated Emission Cross-Section

Stimulated transitions of rare-earth ions in glass tend to be predominately driven by electric dipole interactions [43.65]. For a given transition of interest, the spontaneous emission probability can be expressed in cgs units as [43.6, 85]:

$$A = \frac{64\pi^4 e^2 \chi}{3h (2J+1) \lambda_p^3} S \quad (43.10)$$

where J is the total angular momentum of the upper lasing level, λ_p is the peak emission wavelength, and χ is the local field correction factor. For electric-dipole interactions of an ion in a dielectric medium, $\chi \approx n(n^2 + 2)^2/9$, with n the refractive index of the host glass. S is the quantum-mechanical line strength for the transition. Further, the peak stimulated emission cross-section can be expressed in terms of the spontaneous emission probability:

$$\sigma_p = \left(\frac{\lambda_p^4}{8\pi c n^2 \Delta\lambda_{\text{eff}}} \right) A \quad (43.11)$$

where $\Delta\lambda_{\text{eff}}$ is the effective linewidth of the transition. From (43.10) and (43.11), the host-dependent factors that influence σ_p are the refractive index, the line strength, the effective linewidth, and to a lesser extent the peak emission wavelength (which typically varies only slightly between different hosts). The local field correction factor is significant in hosts with large refractive

index, and can result in an enhancement of the stimulated emission cross-section and a reduction of the radiative lifetime. This is especially true for chalcogenide glass hosts, which typically have refractive index in the 2 to 3 range.

Metastable Lifetime

In the limit of low rare-earth dopant concentration, W_{ET} in (43.9) is zero because the ions are sufficiently well separated to negate their interaction. If the difference in energy between the upper lasing level and the adjacent state is several times the effective phonon energy, then W_{MP} can be neglected to first order. Further neglecting impurity quenching, we can then assert that $\tau_0 \approx 1/A$, where τ_0 is the metastable lifetime in the limit of low rare-earth concentration. From Einstein's relations, $A \approx \sigma_p$, so it follows that $\tau_0 \approx 1/\sigma_p$. In short, hosts that result in an enhancement of the peak stimulated emission cross-section (due to an enhancement of the electric-dipole interaction or because of a high local field correction factor) will also result in a reduction in metastable lifetime. In other words, both stimulated and spontaneous emission rates are enhanced.

The inverse scaling of stimulated emission cross-section and metastable lifetime represents a tradeoff, as it is generally desirable for the lifetime to be as large as possible. Some of the advantages of long lifetime were discussed in Sect. 43.3.1. In addition, the pumping efficiency (gain per applied pump power) of a waveguide amplifier scales directly with τ [43.6, 81] and, therefore, the threshold for CW lasing scales inversely with τ [43.86].

Concentration Quenching

Glasses differ greatly in the amount of a given rare-earth dopant that they are able to dissolve. To avoid problematic ion-ion interactions, the rare-earth ions should be

Table 43.1 Representative parameters for erbium ions embedded in various types of glass (after [43.61, 64, 65])

Glass host	Refractive index n	Peak stimulated emission cross-section σ_p (10^{-21} cm^2)	Metastable lifetime τ (ms)	Quenching concentration ρ_q (10^{20} cm^{-3})	Effective luminescence bandwidth (nm)
Silica	1.46	7	12	—	11
Amorphous Al_2O_3	1.64	6	7.8	—	55
Aluminosilicate	1.5	5.7	10	3.9–6.0	43
Phosphate	1.56	8	10	3.9–8.6	27
Fluoride	1.53	5	9	3.8–5.3	63
Tellurite	2.1	13	3.3	—	80
Sulfide	2.4	20	2.5	3.2	—

incorporated homogeneously into the glass structure. In the extreme case of high concentration, the rare-earth ions will form microscopic clusters (phase separation). Such clustering is highly detrimental, as typically all of the ions within a cluster are effectively removed from the desired stimulated emission process [43.81]. As is well known [43.64, 65], the onset of clustering occurs at quite small values (≈ 0.1 at %) in pure silica glass. The addition of Al_2O_3 to SiO_2 allows silica-based fibers to dissolve a significantly higher concentration. For example, a value of 10–20 for the Al ion to Er ion concentration ratio has been shown to greatly reduce clustering of Er ions in silica glass [43.86]. Interestingly, the addition of Ga (with similar ratio) to chalcogenide glasses has been shown to provide a similar reduction in rare-earth ion clustering [43.41]. These additives essentially modify the glass network, and create sites for isolated rare-earth ions to be incorporated. For similar reasons, multicomponent aluminosilicate and phosphosilicate glasses have been favored in the development of integrated waveguide amplifiers, where the rare-earth concentration must be orders of magnitude higher than in fibers [43.81].

Even in the absence of significant clustering, ion–ion interactions can occur at high concentrations. This is simply due to the reduction in inter-ion spacing, and is exacerbated by any non-uniform (non-homogeneous) distribution of the rare-earth ions in the glass host. These interactions are manifested by a reduction in the metastable lifetime, often well described by the semi-empirical expression [43.61, 65]:

$$\tau(\rho) = \frac{\tau_0}{1 + (\rho/\rho_Q)^p}$$

(43.12)

where ρ is the rare-earth ion concentration, ρ_Q is the so-called quenching concentration, and p is a fitting parameter ($p \approx 2$ for interactions between pairs of ions). The parameter ρ_Q is useful for comparing glasses in terms of their ability to uniformly dissolve a given ion.

Phonon Energies

The characteristic phonon energies of a glass depend on the weights of its constituent atoms and the strength and nature (ionic or covalent) of its bonds. Typical values are shown in Table 43.2. The rate of multi-phonon decay (WMP) between two energy levels depends exponentially on the number of phonons required to bridge the energy gap. Thus, the phonon energy has a great impact on the ultimate efficiency of a desired radiative transition. Low phonon energy can be a good or bad

thing, depending on the transition of interest and the particular pumping scheme.

Mid- to far-infrared transitions of rare-earth ions can exhibit reasonably high quantum efficiency in low phonon energy hosts, such as fluoride, tellurite, and especially chalcogenide glasses. If the same ions are embedded in silicate or phosphate glasses, these transitions are completely quenched by non-radiative processes at room temperature. For this reason, rare-earth doped chalcogenide glasses are of interest for realization of long wavelength amplifiers and lasers [43.41, 64]. Also unique to low phonon energy hosts is the possibility of efficient upconversion lasers [43.87]. In simple terms, the long lifetimes of numerous energy levels allows processes such as ion–ion interactions and excited state absorption (ESA) to efficiently populate the higher energy levels. By contrast, these levels are rapidly depopulated by phonons in oxide glasses. On the other hand, population of the higher levels is highly detrimental if the desired transition is between two lower levels. For example, the efficient pumping of EDFAs using 980 nm wavelength sources relies on the rapid decay (via multiphonon processes) of ions from the $^4\text{I}_{11/2}$ pumping level to the $^4\text{I}_{13/2}$ lasing level. In fluoride and chalcogenide glass hosts, ions raised above the $^4\text{I}_{11/2}$ level tend to become trapped in higher levels (so-called ‘population bottlenecking’ [43.65]). Cerium co-doping has been shown to alleviate this problem [43.66]. Another approach is addition of light elements to the glass network, to increase the phonon energy [43.79].

43.3.3 Progress Towards Integrated Light Sources in Glass

Per the preceding discussion, important goals for glass-based lasers include size reduction and the need for simplified optical or (ideally) electrical pumping

Table 43.2 Characteristic maximum phonon energies for a variety of glass hosts (after [43.6, 65])

Glass host	Phonon energy (cm ⁻¹)
Borate	1400
Phosphate	1200
Silicate	1100
Germanate	900
Tellurite	700
Heavy-metal fluoride	500
Chalcogenide (sulfide)	450
Chalcogenide (selenide)	350

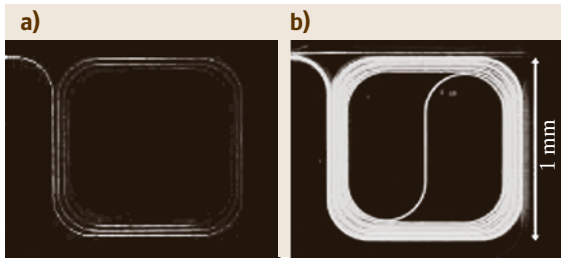


Fig. 43.8 Spiral geometry demonstrated for Al_2O_3 -based waveguide amplifiers, where erbium was introduced by ion implantation (*left*) and co-sputtering (*right*). The bright green emission is a result of ion–ion interactions causing up-conversion for the co-sputtered sample [43.89]

schemes. It should be noted that integration brings its own advantages. As is well known, waveguide lasers have greatly enhanced pump efficiency (gain per applied pump power) relative to their bulk solid-state counterparts. This is due to the confinement of the pump beam and the stimulated emission beam within a small volume. As explored recently in [43.88], this benefit scales with confinement. Amplifiers based on very high index contrast waveguides can have extremely high pump efficiency, as well as requiring only a small amount of wafer real estate. Thus, while the gain per unit length is limited by concentration quenching in rare-earth doped glasses, devices can still be very compact through the use of folded spiral waveguide layouts as illustrated in Fig. 43.8 [43.89].

In terms of maximizing gain per unit length, phosphate glasses have produced the best results to date. This is due to their ability to uniformly dissolve large concentrations of rare-earth ions, and the high cross-sections exhibited by those ions. The high phonon energy can also be an advantage as discussed above. Phosphate glass disks (typically ≈ 1 mm thick) have been widely used for realization of compact microlasers [43.90] and ultrafast modelocked lasers with pulse length as short as 220 fs [43.91] and repetition rate exceed-

ing 40 GHz [43.92]. Phosphate glass fiber lasers with a length of only 7 cm have recently generated 9.3 W of output power [43.93]. Sputtered phosphate glass waveguides have produced gains exceeding 4 dB/cm [43.94]. More recently, a phosphate glass co-doped with 8 at % erbium and 12 at % ytterbium was used to realize ion-exchange waveguides exhibiting 4.1 dB gain in only 3 mm length [43.95]. At such high concentrations, the rare-earth constituents can hardly be called dopants as they exceed the concentration of some of the host glass constituents.

On the theme of simplified pumping options, so-called broadband sensitizers [43.96] are of great interest. A broadband sensitizer is essentially a co-dopant that is able to efficiently absorb pump energy over a wide wavelength range and subsequently excite nearby rare-earth ions to the desired lasing level. This might enable planar waveguide amplifiers that are pumped transversely using low cost broadband light sources [43.64]. Silicon nanoclusters (embedded in oxide glass) are receiving the most attention at present. In addition to acting as a broadband sensitizer, there is evidence that these nanoclusters can enhance the stimulated emission cross-section of co-doped rare-earth ions by 1–2 orders of magnitude. Han et al. [43.97] have reported a waveguide amplifier (on a silicon platform) exhibiting signal enhancement of 7 dB/cm, in spite of a very low erbium concentration. The same group has also reported gain using transverse pumping by LED arrays [43.98]. Finally, the presence of a large concentration of silicon clusters makes electrical pumping a realistic possibility for some silicon-rich oxide glasses [43.64].

It seems very likely that chip-scale glass-based lasers are on the verge of playing a central role in photonics. These lasers will deliver moderate power in a very small package, and will offer extreme spectral purity or stable, high-repetition rate pulses. They might even satisfy the long-standing desire for a silicon-based laser [43.64, 91], facilitating the convergence of photonics and electronics.

43.4 Summary

Glasses have numerous advantages (ease of component fabrication, low scattering loss, flexible atomic structure) that have made them the workhorse material for traditional bulk optics applications. These properties can provide equal or greater advantage in the realization of microphotonic devices and circuits. The glass transi-

tion can enable manufacturing techniques unavailable to other materials, and realization of devices having surface roughness approaching fundamental limits. Owing to this, glasses underlie the highest Q microcavities and some of the lowest loss photonic crystal and microphotonic waveguides reported to date. The metastability

of glass enables a rich array of processing and post-processing options. Direct patterning of waveguides, gratings, and other microphotonic elements by energetic beams is widely studied. These methods of material modification are also promising for post-fabrication trimming of devices. On the other hand, metastability has implications for processing and aging of glass devices.

Rare-earth doped glasses offer numerous advantages relative to other solid-state laser media, especially for realization of ultra-low noise single frequency and ultrafast lasers. Recent advances have greatly increased the maximum gain per unit length, and point to the potential for compact, on-chip glass-based light sources. Progress with respect to 2nd and 3rd order nonlinear effects in glasses is ongoing. It is expected that cm-scale switching or pulse shaping devices based on glasses will become feasible, at least for niche applications.

Overall, it is clear that glasses can contribute greatly to the development of compact, low-loss, multifunction optics integrated with electronics.

Defining Terms

Amorphous Metaloxides are glassy alloys of a transition metal with oxygen, typical examples being TiO_2 , Ta_2O_5 , Nb_2O_5 , and Y_2O_3 . In bulk form, these materials are typically polycrystalline or crystalline ceramics. However, amorphous thin films can be deposited with relative ease, and they have been widely used as high index layers in optical filter design and as dielectric layers in the microelectronics industry.

Broadband Sensitizer is typically some species co-doped along with rare-earth ions into a glass host, in order to increase the pumping efficiency or radiative efficiency of the rare-earth ions. Various sensitizers have been demonstrated, including silicon nanoclusters, silver ions, and other rare earth ions (such as in the sensitization of erbium by ytterbium).

Chalcogenide Glasses are amorphous alloys containing S, Se, and/or Te. Typical examples include Se, GeS_2 , GeSe_2 , As_2S_3 , As_2Se_3 , and As_2Te_3 . By intermixing these and other binary chalcogenide glasses, a wide variety of multicomponent glasses can be formed. Further, a wide range of non-stoichiometric compositions is possible. Several compositions have become standard industrial materials, including $\text{Ge}_{33}\text{As}_{12}\text{Se}_{55}$ and $\text{Ge}_{28}\text{Sb}_{12}\text{Se}_{60}$. The chalcogenide glasses are characterized by narrow bandgaps and good transparency in the mid to far infrared wavelength range.

Concentration Quenching refers to the reduction in luminescence efficiency and luminescence lifetime of a laser glass when the rare-earth dopant concentration is high. Quenching is due to interactions between closely spaced rare-earth ions at high concentrations. These interactions create new pathways, other than the desired radiative decay, for the ions to relax to the ground state after they have been raised to a desired lasing level by pump energy.

Devitrification refers to the transition of a glassy material to its lower energy crystalline state. This process is usually driven by thermal energy, such as if the material is held at some characteristic temperature above its glass transition temperature. The difference between the crystallization temperature and the glass transition temperature for a particular glass is one measure of its stability.

Fluoride Glasses are multicomponent glasses, typically based on fluorides of zirconium, barium, lead, gallium, lanthanum, aluminium, and sodium. They have a wide transparency range, from ultraviolet to mid-infrared wavelengths. They also have low characteristic phonon energies and can dissolve large concentrations of rare-earth ions. For these reason, they are extremely popular as hosts for rare-earth doped amplifiers and lasers operating in the UV-vis and mid-infrared regions.

Glass Transition Temperature is the approximate temperature at which a material changes from a supercooled liquid to an amorphous solid, or vice versa. The transition is marked by an abrupt but continuous change in slope of the specific volume and enthalpy versus temperature curves. Viscosity varies rapidly near the glass transition temperature, which is also sometimes called the softening temperature.

High Index Contrast refers to waveguides or devices fabricated using two or more materials that have very different refractive index. High index contrast is the basis for the confinement of light to very small cross-sectional area waveguides or very small volume optical cavities, either using total internal reflection or photonic bandgap effects. High index contrast thus is the basis for increased density of optical integrated circuits.

Integrated Optics/Photonics refers to the manufacture of photonic elements and circuits on a planar substrate, typically using thin film deposition, lithography, and etching steps. Typically, the substrate is a glass or semi-

conductor wafer and the photonic elements are guided wave devices.

Metastability is a term that refers to the non-equilibrium nature of glasses or amorphous solids. Amorphous solids have excess internal energy relative to the corresponding crystalline state or states of the same material. The method of manufacture, such as melt quenching, inhibits a transition to the lowest energy crystalline state.

Microphotonics refers to the chip-scale manufacture of optical and photonic waveguide circuitry, using processing techniques borrowed from the microelectronics industry. Related to this is the need for high-density integrated optics, as facilitated by high index contrast waveguides and photonic crystals. By usual definition, microphotonics refers specifically to the monolithic manufacture of optical and photonic elements on silicon (CMOS) chips.

Photoinduced Effects are changes in the properties of a glass induced by light, involving transitions between metastable states of the glass or changes in defect sites within the glass. Typically, a laser beam is used to locally modify the refractive index, density, absorption coefficient, etc. of the glass. These processes are widely used to pattern photonic structures such as Bragg gratings, waveguides, and refractive lenses into glasses.

Planar Lightwave Circuit or PLC refers to the industrially established processes for manufacturing integrated

optics devices in silica-based glasses deposited on silicon wafers. Typically, the glass layers are deposited by chemical vapor deposition or flame hydrolysis. These technologies were developed mainly for applications in fiber optics, and are widely used to manufacture wavelength multiplexers.

Reflow is the process of heating a glass above its glass transition temperature, to the point that its viscosity is sufficiently reduced to enable the material to flow. In combination with surface tension effects or other external forces, reflow is often exploited in the reshaping of optical devices.

Supercooling Temperature is the difference between the glass transition temperature and the in-use temperature for a glass-based device. For a large (small) supercooling temperature, the structural relaxation rate is low (high).

Structural Relaxation is essentially an aging effect associated with glasses. Because glasses are metastable materials with random network structures, they are inherently subject to short or long term changes in material properties. Often, structural relaxation is manifested by a change in specific volume (densification) at fixed temperature versus time. The rate of such changes is extremely sensitive to the difference between the glass transition temperature and the observation temperature. Structural relaxation can be induced rapidly by an annealing step, in which the glass is heated near its glass transition temperature for some period of time.

References

- 43.1 H. Rawson: Glass and its History of Service, Part A, IEE Proceedings **135**(6), 325–345 (1988)
- 43.2 W. J. Tropf, M. E. Thomas, T. J. Harris: *OSA Handbook of Optics, Vol. II*, 2nd edn. (McGraw-Hill, New York 1995)
- 43.3 K. Hirao, T. Mitsuyu, J. Si, J. Qiu: *Active Glass for Photonic Devices, Photoinduced Structures and Their Application* (Springer, Berlin, Heidelberg 2001)
- 43.4 Z. Yin, Y. Jaluria: Neck-down and thermally induced defects in high-speed optical fiber drawing, J. Heat Transfer **122**(2), 351–362 (2000)
- 43.5 W. A. Gambling: IEEE J. Sel. Top. Quant. Elec. **6**, 1084 (2000)
- 43.6 M. Yamane, Y. Asahara: *Glasses for Photonics* (Cambridge Univ. Press, Cambridge 2000)
- 43.7 J. Lucas: Curr. Op. Sol. St. Mat. Sci. **4**, 181 (1999)
- 43.8 H. Ma, A. K.-Y. Jen, L. R. Dalton: Adv. Mat. **14**, 1339 (2002)
- 43.9 M. R. Poulsen, P. I. Borel, J. Fage-Pederson, J. Hubner, M. Kristensen, J. H. Povlsen, K. Rottwitt, M. Svalgaard, W. Svendsen: Opt. Eng. **42**, 2821 (2003)
- 43.10 K. Okamoto: *Integrated Optical Circuits and Components, Design and Applications* (Dekker, New York 1999), Chapt. 4
- 43.11 S. Torquato: Nature **405**, 521 (2000)
- 43.12 P. G. Debenedetti, F. H. Stillinger: Nature **410**, 259 (2001)
- 43.13 C. A. Angell, K. L. Ngai, G. B. McKenna, P. F. McMillan, S. W. Martin: J. Appl. Phys. **88**, 3113 (2000)
- 43.14 J. M. Saiter, M. Arnoult, J. Grenet: Phys. B: Cond. Matt. **355**, 370 (2005)
- 43.15 J. C. Anderson, K. D. Leaver, R. D. Rawlings, J. M. Alexander: *Materials Science*, 4th edn. (Chapman Hall, London 1990)
- 43.16 B. Hendriks, S. Kuiper: IEEE Spectrum **41**, 32 (2004)

- 43.17 G. P. Agrawal: *Fiber-Optic Communication Systems*, 2nd edn. (Wiley, New York 1997)
- 43.18 Y. P. Li, C. H. Henry: IEE Proc.-Optoelectron **143**, 263 (1996)
- 43.19 R. R. A. Syms, W. Huang, V. M. Schneider: Elec. Lett. **32**, 1233 (1996)
- 43.20 R. R. A. Syms, A. S. Holmes: IEEE Phot. Tech. Lett. **5**, 1077 (1993)
- 43.21 T. J. Kippenberg, S. M. Spillane, B. Min, K. J. Vahala: IEEE J. Sel. Top. Quant. Elec. **10**, 1219 (2004)
- 43.22 D. W. Vernooy, V. S. Ilchenko, H. Mabuchi, E. W. Streed, H. J. Kimble: Opt. Lett. **23**, 247 (1998)
- 43.23 M. L. Gorodetsky, A. A. Savchenkov, V. S. Ilchenko: Opt. Lett. **21**, 453 (1996)
- 43.24 L. Tong, R. R. Gattass, J. B. Ashcom, S. He, J. Lou, M. Shen, I. Maxwell, E. Mazur: Nature **426**, 816 (2003)
- 43.25 P. J. Roberts, F. Couny, H. Sabert, B. J. Mangan, D. P. Williams, L. Farr, M. W. Mason, A. Tomlinson, T. A. Birks, J. C. Knight, P. St. J. Russell: Opt. Express **13**, 236 (2005)
- 43.26 C.-T. Pan, C.-H. Chien, C.-C. Hsieh: Appl. Opt. **43**, 5939 (2004)
- 43.27 M. He, X.-C. Yuan, N. Q. Ngo, J. Bu, V. Kudryashov: Opt. Lett. **28**, 731 (2003)
- 43.28 M. He, X.-C. Yuan, J. Bu: Opt. Lett. **29**, 2004 (2004)
- 43.29 Y. A. Vlasov, S. J. McNab: Opt. Express **12**, 1622 (2004)
- 43.30 F. Grillot, L. Vivien, S. Laval, D. Pascal, E. Cassan: IEEE Phot. Tech. Lett. **16**, 1661 (2004)
- 43.31 T. Barwicz, H. I. Smith: J. Vac. Sci. Tech. B **21**, 2892 (2003)
- 43.32 D. K. Armani, T. J. Kippenberg, S. M. Spillane, K. J. Vahala: Nature **421**, 925 (2003)
- 43.33 V. Van, P. P. Absil, J. V. Hryniewicz, P.-T. Ho: J. Light. Tech. **19**, 1734 (2001)
- 43.34 K. J. Vahala: Optical Microcavities, Nature **424**, 839–851 (August 2003)
- 43.35 S. J. McNab, N. Moll, Y. A. Vlasov: Opt. Express **11**, 2927 (2003)
- 43.36 P. K. Gupta, D. Inness, C. R. Kurkjian, Q. Zhong: J. Non-Crystalline Sol. **262**, 200 (2000)
- 43.37 D. P. Bulla, W.-T. Li, C. Charles, R. Boswell, A. Ankiewicz, J. Love: Appl. Opt. **43**, 2978 (2004)
- 43.38 X. H. Zhang, Y. Guimond, Y. Bellec: J. Non-crystalline Sol. **326&327**, 519 (2003)
- 43.39 A. K. Mairaj, X. Feng, D. P. Shepherd, D. W. Hewak: Appl. Phys. Lett. **85**, 2727 (2004)
- 43.40 A. K. Mairaj, R. J. Curry, D. W. Hewak: Appl. Phys. Lett. **86**, 094102 (2005)
- 43.41 J. S. Sanghera, L. B. Shaw, I. D. Aggarwal: *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2nd edn. (Dekker, New York 2001), Chapter 9
- 43.42 Z. Zhang, G. Xiao, C. P. Grover: Appl. Opt. **43**, 2325 (2004)
- 43.43 P. K. Tien: Appl. Opt. **10**, 2395 (1971)
- 43.44 R. Ulrich: J. Vac. Sci. Tech. **11**, 156 (1974)
- 43.45 P. K. Tien, A. A. Ballman: J. Vac. Sci. Tech. **12**, 892 (1974)
- 43.46 J. M. Mir, J. A. Agostinelli: J. Vac. Sci. Tech. A **12**, 1439 (1994)
- 43.47 J. Mueller, M. Mahnke, G. Schoer, S. Wiechmann: AIP Conference Proceedings **709**, 268 (2004)
- 43.48 D. Cangialosi, M. Wubbenhorst, H. Schut, A. van Veen, S. J. Picken: Phys. Rev. B **69**, 134206–1 (2004)
- 43.49 M. B. J. Diemeer: AIP Conference Proceedings **709**, 252 (2004)
- 43.50 N. Daldosso, M. Melchiorri, F. Riboli, F. Sbrana, L. Pavesi, G. Pucker, C. Kompocholis, M. Crivellari, P. Bellutti, A. Lui: Mat. Sci. Semicond. Proc. **7**, 453 (2004)
- 43.51 A. V. Kolobov (Ed.): *Photo-induced Metastability in Amorphous Semiconductors* (Wiley-VCH, Weinheim 2003)
- 43.52 K. Tanaka: C.R. Chimie **5**, 805 (2002)
- 43.53 H. Haeiwa, T. Naganawa, Y. Kokubun: IEEE Phot. Tech. Lett. **16**, 135 (2004)
- 43.54 K. Minoshima, A. M. Kowalevich, E. P. Ippen, J. G. Fujimoto: Opt. Express **10**, 645 (2002)
- 43.55 N. Ponnampalam, R. G. DeCorby, H. T. Nguyen, P. K. Dwivedi, C. J. Haugen, J. N. McMullin, S. O. Kasap: Opt. Express **12**, 6270 (2004)
- 43.56 K. O. Hill: IEEE J. Sel. Top. Quant. Elec. **6**, 1186 (2000)
- 43.57 P. R. Herman, R. S. Marjoribanks, A. Oettl, K. Chen, I. Konovalov, S. Ness: Appl. Surf. Sci. **154–155**, 577 (2000)
- 43.58 M. Aslund, J. Canning: Opt. Lett. **25**, 692 (2000)
- 43.59 C. Florea, K. A. Winick: J. Light. Tech. **21**, 246 (2003)
- 43.60 A. Zakery, Y. Ruan, A. V. Rode, M. Samoc, B. Luther-Davies: J. Opt. Soc. Am. B. **20**, 1844 (2003)
- 43.61 X. Orignac, D. Barbier, X. Min Du, R. M. Almeida, O. McCarthy, E. Yeatman: Opt. Mat. **12**, 1 (1999)
- 43.62 R. A. Bellman, G. Bourdon, G. Alibert, A. Beguin, E. Guiot, L. B. Simpson, P. Lehuède, L. Guiziou, E. LeGuen: J. Electrochem. Soc. **151**, 541 (2004)
- 43.63 G.-L. Bona, R. Germann, B. J. Offrein: IBM J. Res. & Dev. **47**, 239 (2003)
- 43.64 A. J. Kenyon: Prog. Quant. Elec. **26**, 225 (2002)
- 43.65 M. J. Miniscalco: *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2nd edn. (Dekker, New York 2001), Chapt. 2
- 43.66 Y. Gao, B. Boulard, M. Couchaud, I. Vasilief, S. Guy, C. Duverger, B. Jacquier: Opt. Mat. **27**, 195–199 (2005)
- 43.67 T. Barwicz, M. A. Popovic, P. T. Rakich, M. R. Watts, H. A. Haus, E. P. Ippen, H. I. Smith: Opt. Express **12**, 1437 (2004)
- 43.68 T. Kawashima, K. Miura, T. Sato, S. Kawakami: Appl. Phys. Lett. **77**, 2613 (2000)
- 43.69 R. Rabaday, I. Avrutsky: Appl. Opt. **44**, 378 (2005)
- 43.70 H. Hirota, M. Itoh, M. Oguma, Y. Hibino: IEEE Phot. Tech. Lett. **17**, 375 (2005)
- 43.71 C.-Y. Tai, J. S. Wilkinson, N. M. B. Perney, M. Cateirina Netti, F. Cattaneo, C. E. Finlayson, J. J. Baumberg: Opt. Express **12**, 5110 (2004)

- 43.72 B. Unal, C.-Y. Tai, D. P. Shepherd, J. S. Wilkin-son, N. M. B. Perney, M. Caterina Netti, G. J. Parker: Nd:Ta2O5 rib waveguide lasers, *Appl. Phys. Lett.* **86**, 021110 (2005)
- 43.73 Y. Kokubun, Y. Hatakeyama, M. Ogata, S. Suzuki, N. Zaizen: *IEEE J. Sel. Top. Quant. Elec.* **11**, 4 (2005)
- 43.74 T. Sato, K. Miura, N. Ishino, Y. Ohtera, T. Tamamura, S. Kawakami: *Opt. Quant. Elec.* **34**, 63 (2002)
- 43.75 D. R. MacFarlane: *Ceramics International* **22**, 535 (1996)
- 43.76 A. Feigel, M. Veinger, B. Sfez, A. Arsh, M. Klebanov, V. Lyubin: *Appl. Phys. Lett.* **83**, 4480 (2003)
- 43.77 G. Lenz, S. Spalter: *Nonlinear Photonic Crystals* (Springer, Berlin, Heidelberg 2003), Chapt. 11
- 43.78 K. Ogusu, J. Yamasaki, S. Maeda, M. Kitao, M. Minakata: *Opt. Lett.* **29**, 265 (2004)
- 43.79 S. Hocde, S. Jiang, X. Peng, N. Peyghambarian, T. Luo, M. Morrell: *Opt. Mat.* **25**, 149 (2004)
- 43.80 R. Nayak, V. Gupta, A. L. Dawar, K. Sreenivas: *Thin Sol. Films* **445**, 118 (2003)
- 43.81 D. Barbier: *Integrated Optical Circuits and Components, Design and Applications* (Dekker, New York 1999), Chapt. 5
- 43.82 B. E. Callicoatt, J. B. Schlager, R. K. Hickernell, R. P. Mirin, N. A. Sanford: *IEEE Circuits & Devices Mag.* **19**, 18 (September 2003)
- 43.83 E. DeSurvire: *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2ed. (Dekker, New York 2001), Chapt. 10
- 43.84 A. Yariv: *Optical Electronics in Modern Communications*, 5th edn. (Oxford Univ. Press, New York 1997)
- 43.85 Q. Wang, N. K. Dutta: *J. Appl. Phys.* **95**, 4025 (2004)
- 43.86 M. J. F. Digonnet: *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2nd edn. (Dekker, New York 2001), Chapt. 3
- 43.87 D. S. Funk, J. G. Eden: *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2nd edn. (Dekker, New York 2001), Chapt. 4
- 43.88 S. Saini, J. Michel, L. C. Kimerling: *J. Light. Tech.* **21**, 2368 (2003)
- 43.89 P. G. Kik, A. Polman: *J. Appl. Phys.* **93**, 5008 (2003)
- 43.90 P. Laporta, S. Taccheo, S. Longhi, O. Svelto, C. Svelto: *Opt. Mat.* **11**, 269 (1999)
- 43.91 F. J. Grawert, J. T. Gopinath, F. O. Ilday, H. M. Shen, E. P. Ippen, F. X. Kartner, S. Akiyama, J. Liu, K. Wada, L. C. Kimerling: *Opt. Lett.* **30**, 329 (2005)
- 43.92 U. Keller: *Nature* **424**, 831 (2003)
- 43.93 T. Qiu, L. Li, A. Schulzgen, V. L. Temyanko, T. Luo, S. Jiang, A. Mafey, J. V. Moloney, N. Peyghambarian: *IEEE Phot. Tech. Lett.* **16**, 2592 (2004)
- 43.94 Y. C. Yan, A. J. Faber, H. de Waal, P. G. Kik, A. Polman: *Appl. Phys. Lett.* **71**, 2922 (1997)
- 43.95 F. D. Patel, S. DiCarolis, P. Lum, S. Venkatesh, J. N. Miller: *IEEE Phot. Tech. Lett.* **16**, 2607 (2004)
- 43.96 A. Polman, F. C. J. M. van Veggel: *J. Opt. Soc. Am. B* **21**, 871 (2004)
- 43.97 H.-S. Han, S.-Y. Seo, J. H. Shin, N. Park: *Appl. Phys. Lett.* **81**, 3720 (2002)
- 43.98 J. Lee, J. H. Shin, N. Park: *J. Light. Tech.* **23**, 19 (2005)

44. Optical Nonlinearity in Photonic Glasses

A brief review of optical nonlinearity in photonic glasses is given. For third-order nonlinearity, the relationship between two-photon absorption and nonlinear refractive index is considered using a formalism developed for crystalline semiconductors. Stimulated light scattering and supercontinuum generation in optical fibers are also introduced. Prominent resonant-type nonlinearity in particle-embedded glasses is described. For second-order nonlinearity, a variety of poling methods are summarized. Finally, it is pointed out that various photoinduced changes can appear when excited by linear and nonlinear optical processes, and this is related to glass structure.

44.1 Third-Order Nonlinearity in Homogeneous Glass	1064
44.1.1 Experimental	1064
44.1.2 Theoretical Treatment	1065
44.1.3 Stimulated Light Scattering and Supercontinuum Generation	1068
44.2 Second-Order Nonlinearity in Poled Glass	1069
44.3 Particle-Embedded Systems	1070
44.4 Photoinduced Phenomena	1071
44.5 Summary	1072
References	1072

New developments in optical fibers and pulsed lasers have prompted increasing interest in optical nonlinearity in photonic glasses [44.1–3]. Third-order polarization yields several nonlinear phenomena, such as intensity-dependent absorption and intensity-dependent refractive index, which can be utilized in power stabilizers and all-optical switches. On the other hand, in high-power glass lasers, self-focusing effects arising from intensity-dependent increases in refractive index pose serious problems. Then again, the second-order polarization that occurs in poled glasses can be utilized in second harmonic generation for example. The present chapter provides a brief review of optical nonlinearity in *inorganic* glasses. At this point we should mention that, in many respects, organic polymers exhibit similar features to those of glass [44.4, 5]. In general, glass is more stable, while polymers can provide greater nonlinearity, so these two types of material are competitors in practical applications.

Glass also competes with crystals. A great advantage of glass is the ability to control its structure at three scales. First, the atomic composition of the glass can be tailored continuously. For instance, we can obtain nonlinear optical glass with any refractive index in the range of 1.4–3.2 at a wavelength of $\approx 1 \mu\text{m}$. Second, the atomic structure can be modified reasonably easily using, e.g., light beams. Such modifications may be regarded as photoinduced phenomena, which can be

employed in order to add second-order nonlinearity to selected regions and so forth. Lastly, macroscopic shapes can be changed into arbitrary bulk forms, fibers, thin layers and microparticles [44.6, 7].

Here, fiber and film waveguides may be very important for nonlinear applications for two reasons. One is that the light power density can be increased by reducing the lateral size, i. e. the film thickness or fiber diameter, to submicron scales. The other is, that fibers can provide long lengths for light–glass interactions that are not limited by diffraction. These scale factors produce apparent enhancements in nonlinear effects in glasses, although the intrinsic nonlinearity may be smaller than that in crystals.

Before proceeding further, it would be useful to introduce a nonlinear formula [44.4, 5]. For simplicity, we take the polarization P and the electric field E to be scalar quantities. Then, very simply, P can be written down in the cgs units as

$$P = \chi^{(1)}E + \chi^{(2)}EE + \chi^{(3)}EEE + \dots, \quad (44.1)$$

where the first term $\chi^{(1)}E$ depicts the conventional linear response, and $\chi^{(2)}$ and $\chi^{(3)}$ represent the second- and the third-order nonlinear susceptibilities. $\chi^{(1)}$ is related to the linear refractive index n_0 via $n_0 = \{1 + 4\pi\chi^{(1)}\}^{1/2}$. On the other hand, the second term provides such time dependence as $\sin\{(\omega \pm 2\omega)t\}$, so that it could produce dc and second-overtone (2ω) signals. In a similar way, the

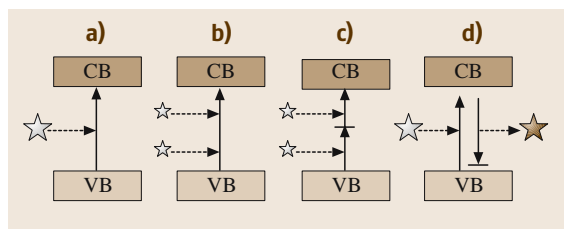


Fig. 44.1a–d Schematic illustrations of (a) one-photon absorption, (b) two-photon absorption, (c) two-step absorption via a mid-gap state, and (d) a Raman-scattering process in a semiconductor with a valence band (VB) and a conduction band (CB)

third could modify the fundamental (ω) signal and generate a third-overtone (3ω). Nevertheless, since a refractive index changes with frequency, it is difficult in the overtone generations to satisfy the so-called phase-matching condition [44.4, 5], e.g. $\varphi_{3\omega} = 3\varphi_{\omega}$, where φ is the phase of electric fields in glasses. The, the fundamental-signal processing may be the most important. Microscopically, the nonlinear terms arise through several mechanisms, such as electronic, atomic (including molecular motions), electrostatic and thermal processes. Among these, electronic processes can provide the fastest responses, with fs–ps time scales, which will be needed for optical information technologies. Accordingly, we will focus on such processes from this point onwards.

The chapter is divided up as follows. In Sect. 44.1, we treat conventional glasses. These are optically isotropic, which gives $\chi^{(2)} = 0$, because the isotropic structure appears to be centrosymmetric. In this case, the lowest nonlinear term becomes $\chi^{(3)}EEE$. When $\chi^{(3)}EEE$ can be written as $\chi^{(3)}IE$, where I is the light intensity, we can define the refractive index n as $n = n_0 + n_2I$, where n_2 is sometimes called the second-order index of refraction [44.5]. In nonlinear optics, this notation poses problems due to the definitions of several parameters. Therefore, the con-

version $n_2(\text{cm}^2/\text{W}) \approx 0.04\chi^{(3)}(\text{esu})/n_0^2$ is frequently employed [44.5]. On the other hand, for light absorption, which is proportional to $\langle E dP/dt \rangle$, where $\langle \dots \rangle$ represents a time average, we need to take the imaginary parts of $\chi^{(1)}$ and $\chi^{(3)}$, which cause one- and two-photon absorption, into account (Fig. 44.1). In such cases, an effective absorption coefficient can be written as $\alpha(\text{cm}^{-1}) + \beta(\text{cm}/\text{W})I(\text{W}/\text{cm}^2)$; n_2 and β will be connected through nonlinear Kramers–Kronig relations.

Section 44.2 focuses upon noncentrosymmetry. From the mid-1980s onwards, several kinds of poling treatments have been found to add $\chi^{(2)}$ to glasses [44.8]. For instance, Österberg and Margulis [44.9] demonstrated intense second harmonic generation in laser-irradiated glass fibers. Such work began with silica, and has more recently been directed towards more complicated glasses. The magnitude of $\chi^{(2)}$ seen in this case may be smaller than those observed in crystals, due to the disordered atomic structures in glasses, but the structural controllability may offer some advantages.

Section 44.3 introduces glasses incorporating fine semiconductor and metal particles. Such nanostructured glasses are known to exhibit large optical nonlinearities [44.4, 5, 10]. However, as will be described, their wider features remain to be studied.

In Sect. 44.4 we consider photoinduced phenomena. The intense light needed to produce optical nonlinearity is likely to modify the structure of glass. Light that is even more intense may cause damage. While this damage can be useful for optical engraving, the structural modification may ultimately be more important. Actually, fiber Bragg gratings [44.1, 11] have been produced commercially using excimer lasers, and nonlinear optical excitation may play an important role in these. Three-dimensional modifications are also produced through controlled light focusing [44.12]. Section 44.5 provides a summary of the chapter. This chapter is based on author's recent review [44.13].

44.1 Third-Order Nonlinearity in Homogeneous Glass

44.1.1 Experimental

Substantial data are available for n_2 at transparent wavelengths [44.4]. However, to the author's knowledge, all such experiments have utilized lasers, and no spectral dependence has been reported. We should also note that, in comparison with n_0 measurements, n_2 evaluations are much more difficult [44.4]. The biggest

difficulty, which is common to all nonlinear experiments, is the light intensity. Normally, light is pulsed and focused, producing complex temporal and spatial profiles, which in turn means that intensity evaluations are not straightforward. For instance, if the profile is Gaussian, can we simply use the peak value to evaluate the nonlinearity? Because of such measuring difficulties and also some other specific problems inherent to

Table 44.1 Linear (E_g , n_0 , α_0) and nonlinear (n_2 , β , β_{\max}) optical properties and figures of merit ($2\beta\lambda_0/n_2$, n_2/α_0) in some glasses. E_g is the optical bandgap energy [44.12], n_0 is the refractive index [44.12], α_0 is the attenuation coefficient [44.16], n_2 is the intensity-dependent refractive index [44.4, 5, 12], β is the two-photon absorption coefficient [44.16], and β_{\max} is the maximum value. Except for E_g and β_{\max} , the values are evaluated at a wavelength of 1–1.5 μm . BK-7 is a borosilicate glass and SF-59 indicates data for lead-silicate glasses with $\approx 57\text{mol.}\%$ PbO

Glass	E_g (eV)	n_0	α_0 (cm^{-1})	n_2 (10^{-20} m^2/W)	β (cm/GW)	β_{\max} (cm/GW)	$2\beta\lambda_0/n_2$	n_2/α_0 (cm^3/GW)
SiO ₂	10	1.5	10^{-6}	2	$< 10^{-2}$	1	< 10	0.2
BK-7	4	1.5		3				
SF-59	3.8	2.0		30	$< 10^{-1}$	10		
As ₂ S ₃	2.4	2.5	10^{-3}	200	10^{-2}	50	0.1	0.02
BeF ₂	10	1.3		0.8				

glasses, reported n_2 values vary, even for a reference SiO₂ glass, over the range $1\text{--}3 \times 10^{-20} \text{ m}^2/\text{W}$ [44.4]. In other glasses, values that differ by a factor of ≈ 10 have been reported [44.4, 14, 15].

However, we can see a general trend in n_2 in Table 44.1. Halide glasses (BeF₂) have the smallest values, light metal oxides (BK-7) and SiO₂ have similar values, heavy metal oxides such as PbO-SiO₂ (SF-59) have larger values, while the largest belong to the chalcogenides (As₂S₃) [44.4, 10]. The largest value reported so far may be $n_2 = 8 \times 10^{-16} \text{ m}^2/\text{W}$, seen in Ag₂₀As₃₂Se₄₈ at wavelength of 1.05 μm [44.21], which is $\approx 10^4$ times as large as that of SiO₂.

Figure 44.2 shows β spectra for SiO₂, As₂S₃ and two PbO-SiO₂ glasses [44.16]. Bi₂O₃-B₂O₃ glasses (Imanishi et al., [44.22]) exhibit similar features to those of lead glass. We can see that all of the β spectra seem to have maxima at midgap regions, $\hbar\omega \approx E_g/2$. However, this is not an absolute rule, since at $\hbar\omega \geq 2.0 \text{ eV}$ As₂S₃ provides two-step absorption (Fig. 44.1), which masks the two-photon absorption signal [44.20]. We also see that the maximal β ($\approx 10^0 \text{ cm}/\text{GW}$) in SiO₂ is markedly smaller than those belonging to the other glasses. Note that such a marked difference does not exist in maximal α , which is $\approx 10^6 \text{ cm}^{-1}$ at the super-bandgap regions of SiO₂ and As₂S₃. It may be worth mentioning here that two-photon excitation produces photocurrents in some glasses [44.23, 24].

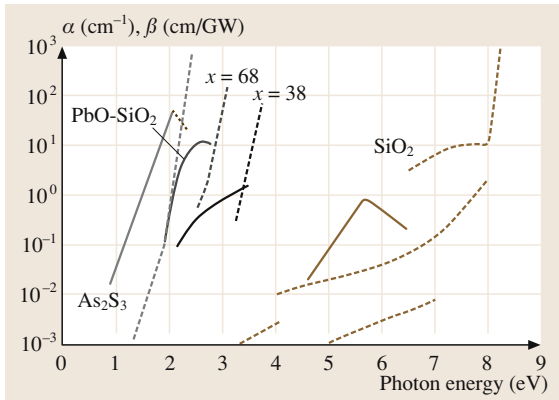


Fig. 44.2 One- (dashed line) and two-photon (solid line) absorption spectra, α and β , in SiO₂ [44.17, 18], 38PbO – 62SiO₂, 68PbO – 32SiO₂ [44.19], and As₂S₃ [44.20]. For SiO₂, in transparent regions ($\hbar\omega \leq 8.0 \text{ eV}$) α should be regarded as an attenuation coefficient; these are not reproducible, probably due to impurities, defects and light scattering. The dotted line for As₂S₃ at $\hbar\omega \geq 2.0 \text{ eV}$ indicates two-step absorption

44.1.2 Theoretical Treatment

For $\chi^{(3)}$, theoretical treatments of glasses, liquids, and crystals are largely the same. Glasses and liquids can be treated similarly in the sense that both are optically isotropic. Hence, many ideas stemming from types of chemical bond pictures have been proposed [44.5]. On the other hand, the magnitude of $\chi^{(3)}$ in glass appears to be similar to that in the corresponding crystal, since their electronic structures are both governed by short-range atomic structures, within $\approx 0.5 \text{ nm}$ [44.25], which is much shorter than the wavelength ($\approx 500 \text{ nm}$) of visible light. Specifically, since n_2 is governed by integrated electronic absorption (44.8), the magnitudes are roughly the same in glass and crystal. For instance, in crystalline and glassy SiO₂ at near-infrared wavelengths, the difference in the n_2 values of both seems to be comparable with their experimental accuracy; i. e. the values for crystalline and glassy SiO₂ are $\approx 1.14 \times 10^{-13}$ and $0.85 \times 10^{-13} \text{ esu}$ respectively [44.4]. Then, using a formula derived for semiconductor crystals, we can apply

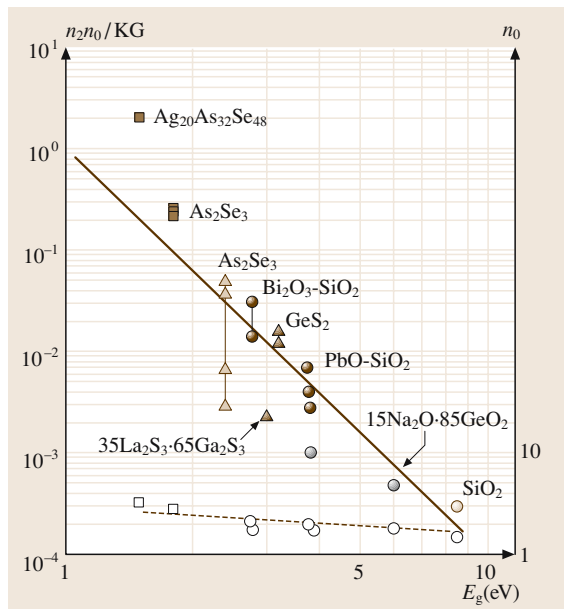


Fig. 44.3 The Sheik–Bahae relation $n_2n_0 = KG(\hbar\omega/E_g)/E_g^4$ (solid line), related data (solid symbols), the Moss rule $n_0^4E_g = 77$ (dashed line), and related data (open symbols) for an oxide (circles), sulfide (triangles) and selenide (squares). The four data for As_2Se_3 are obtained from different publications, while those for PbO-SiO_2 with slightly different E_g correspond to different compositions. The illustration is modified from its previous form [44.16] due to the additions of $\text{Ag}_{20}\text{As}_{32}\text{Se}_{48}$ [44.21], Bi_2O_3 -silica glasses [44.26], $35\text{La}_2\text{S}_3 - 65\text{Ga}_2\text{S}_3$ [44.27], and $15\text{Na}_2\text{O} - 85\text{GeO}_2$ [44.28]

a kind of band picture to glasses [44.16], as described below.

Estimation of Nonlinear Refractivity

Several semi-empirical relationships have been proposed for $\chi^{(3)}$ or n_2 [44.5]. Such relations are useful, since nonlinear optical constants are much more difficult to measure than linear constants. Naturally, since they are simplifications they are also less accurate. The simplest may be the one provided by Wang, $\chi^{(3)} \approx \{\chi^{(1)}\}^4$, which can be regarded as a generalized Miller's rule [44.5]. Boling et al. [44.29] have also derived some relations, among which the simplest may be

$$n_2(10^{-13} \text{ esu}) \approx 391(n_d - 1)/v_d^{5/4}, \quad (44.2)$$

where n_d is the refractive index at the d-line ($\lambda = 588 \text{ nm}$) and v_d is the Abbe number. This relation contains only two macroscopic parameters, which can be

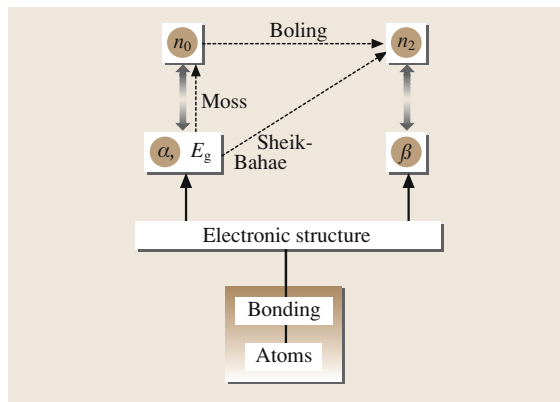


Fig. 44.4 Relationship between atomic structure, electronic structure, linear absorption α including bandgap energy E_g , linear refractive index n_0 , two-photon absorption β , and nonlinear refractive index n_2 . Absorption and refraction are related by the linear and nonlinear Kramers–Kronig relations (double arrows). The Moss rule may be regarded as a simplified Kramers–Kronig relation. The Boling and Sheik–Bahae relations connect n_0 and E_g to n_2 , respectively

easily evaluated, and so it has often been used to estimate n_2 . It actually provides a good approximation for small n_d glasses with $n_d \leq 1.7$ [44.29]. For high n_d glasses, Lines [44.14] proposes a relation that also contains the atomic distance, and others derive more complicated formula [44.30, 31]. Ab initio calculations for TeO_2 -based glasses have also been reported recently [44.32]. However, these relations cannot predict wavelength dependence (Fig. 44.5). The estimated n_2 may be regarded as a long-wavelength limit. In addition, the material is regarded as transparent, so relations between n_2 and β cannot be found.

Nonlinear optical properties of semiconductor crystals have been studied fairly deeply [44.5], and applying the same concepts to glasses appears to be tempting. Tanaka [44.16] has adopted a band picture, which was developed by Sheik-Bahae et al. [44.33], to some glasses. Their concept gives a universal relationship of

$$n_2 = KG(\hbar\omega/E_g)/(n_0E_g^4), \quad (44.3)$$

for crystals with energy gaps of 1–10 eV, where K is a fixed constant and $G(\hbar\omega/E_g)$ is a spectral function (E_g is the bandgap energy). For E_g in glasses, we can take the so-called Tauc gap [44.25] if it is known, or otherwise the photon energy $\hbar\omega$ at $\alpha \approx 10^3 \text{ cm}^{-1}$. As shown in Fig. 44.3, the universal line gives reasonable agreement with published data, while the agreement is

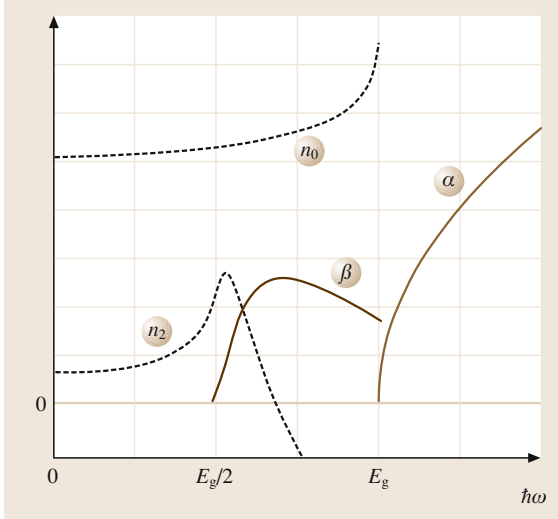


Fig. 44.5 Spectral dependence of linear absorption α , linear refractive index n_0 , two-photon absorption β , and intensity-dependent refractive index n_2 in an ideal amorphous semiconductor with energy gap E_g . Vertical scales are arbitrary

not satisfactory as it is for crystals [44.33]. The poorer agreement for glasses may be partly due to experimental difficulties. In addition, band tail states probably cause larger deviations in smaller bandgap glasses such as $\text{Ag}_{20}\text{As}_{32}\text{Se}_{48}$. Incidentally, as shown by the dashed line in the figure, the Moss rule $n_0^4 E_g = 77$ [44.34], which was originally pointed out for crystalline semiconductors, also gives satisfactory fits for glasses.

Nonlinear Kramers–Kronig Relationship

We can take another approach in order to obtain unified insight into optical nonlinearity (n_2 and β). In this case, we start with absorption and then obtain the refractive index using Kramers–Kronig relationships. As illustrated in Fig. 44.4, this approach is able to provide a complete understanding of atomic structure and optical properties, since absorption spectra can be related to electronic structure more directly.

It is well-known that, for glasses, and neglecting momentum conservation rules, $\alpha(\hbar\omega)$ and $\beta(\hbar\omega)$ can be written as [44.16]

$$\alpha(\hbar\omega) \propto |\langle \phi_f | H | \phi_i \rangle|^2 \int D_f(E + \hbar\omega) D_i(E) dE, \quad (44.4)$$

$$\begin{aligned} \beta(\hbar\omega) &\propto |\Sigma_n \langle \phi_f | H | \phi_n \rangle \langle \phi_n | H | \phi_i \rangle / (E_{ni} - \hbar\omega)|^2 \\ &\times \int D_f(E + 2\hbar\omega) D_i(E) dE, \end{aligned} \quad (44.5)$$

where H is the perturbation Hamiltonian, ϕ is a related electron wavefunction, and D is the density of states, and the subscripts i , n , and f represent the initial, intermediate, and final states. α and β are linked to the electronic structure, i. e. the wavefunctions and densities of states, which is determined by the atomic species and bonding structures involved [44.25]. In this case, $n_0(\omega)$ is expressed using the conventional Kramers–Kronig relation as [44.5]

$$n_0(\omega) = 1 + (c/\pi) \wp \int [\alpha(\Omega)/(\Omega^2 - \omega^2)] d\Omega. \quad (44.6)$$

For a nonlinear response Δn , Hutchings et al. [44.35] have derived the relation

$$\Delta n(\omega; \zeta) = (c/\pi) \wp \int [\Delta\alpha(\Omega; \zeta)/(\Omega^2 - \omega^2)] d\Omega, \quad (44.7)$$

where $\Delta\alpha$ is the nonlinear absorption induced by an excitation at ζ and probed at Ω . Note that this relation is derived from the causality principle for nondegenerate cases, e.g., two-beam experiments with different photon energies, Ω and ζ . However, this relation may also be applied to degenerate cases ($\Omega = \zeta$) as a rough approximation as [44.36]

$$n_2(\omega) = (c/\pi) \wp \int [\beta(\Omega)/(\Omega^2 - \omega^2)] d\Omega. \quad (44.8)$$

Using these formulae, we can roughly predict how β and n_2 depend upon E_g . Equation (44.5) suggests that $\beta \propto 1/E_g^2$, provided that the E_g dependence is governed by $|\Sigma_n 1/(E_{ni} - \hbar\omega)|^2$. This relation may be consistent with the material dependence shown in Fig. 44.2. That is, $\beta_{\max} \propto 1/E_g^2 \approx 1/E_g^3$. Note that this E_g dependence is comparable to $\beta \propto 1/E_g^3$, which is theoretically derived and experimentally confirmed for crystalline semiconductors [44.5, 33]. Then, putting $\beta \propto 1/E_g^2$ into (44.8), and assuming $\int [1/(\Omega^2 - \omega^2)] d\Omega \propto 1/E_g^2$, we obtain $n_2 \propto 1/E_g^4$, which is consistent with (44.3).

Under some plausible assumptions, we can also calculate the spectral dependence of β and n_2 for an ideal amorphous semiconductor, which contains no gap states [44.37]. Figure 44.5 shows that, at $\hbar\omega = E_g/2$, $\beta = 0$ and n_2 is maximal. This is similar to some degree to the feature at $\hbar\omega = E_g$ for α and n_0 . The figure also shows that β becomes maximal at $E_g/2 - E_g$. Note, however, that it is not clear whether we can link this photon energy dependence to the experimental results shown in Fig. 44.2, since the observed spectra appear to have fairly sharp peaks.

Figure of Merit

Several figures of merit have been proposed for evaluating optical devices which utilize n_2 . For instance, Mizrahi et al. [44.40] and others [44.15, 21, 41] emphasize the negative effects of two-photon absorption, and provide the criterion that $2\beta\lambda_0/n_2 < 1$. Then, using the above E_g dependence, we see that $\beta/n_2 \approx E_g^2$, so that small E_g materials are preferable. Actually (as also listed in Table 44.1), in this figure, As_2S_3 appears to be better than SiO_2 . However, the criterion implicitly neglects α . Lines [44.14] uses n_2/α instead. Here, α arises from the so-called residual absorption, which is difficult to estimate quantitatively. In addition, it is the attenuation α_0 instead of the absorption α that could be decisive. Table 44.1 shows that, in this measure, SiO_2 behaves better than As_2S_3 . These figures of merit, however, have presumed only nonresonant electronic contributions with sub-ps responses. More recently, Jha et al. [44.10] utilize $n_2/\tau\alpha$, where τ is the relaxation time. Nevertheless, theoretical predictions of τ remain to be studied.

The most appropriate definition of the figure of merit naturally depends upon the application of interest. For instance, for an optical fiber device, the maximum length might be ≈ 1 m [44.42], since the device must be compact and fast. In this case, the light absorption, which arises from $\alpha + \beta I$, should be suppressed to below 10^{-2} cm^{-1} , or the light propagation loss must be smaller than $\approx 1 \text{ dB/m}$. However, for optical integrated circuits, an effective propagation distance may be $\approx 1 \text{ cm}$, in which the attenuation could be as large as 10^0 cm^{-1} .

We can suggest another idea by taking the spectral dependence shown in Fig. 44.5 into account. That is, the best material that has n_2 at some $\hbar\omega$ is the one which satisfies a bandgap condition of $\hbar\omega = E_g/2$, where α and β are zero and n_2 becomes maximal, provided that there is no gap-state absorption. In practical glass samples, absorption due to impurities, dangling bonds such as E' centers in oxide glasses [44.11, 25], and wrong bonds in chalcogenide glasses [44.25, 37] cannot be neglected. Actually, we can see in Table 44.1 and Fig. 44.2 that residual attenuation exists in nominally transparent regions, parts of which are undoubtedly caused by absorption [44.43]. Note that these mid-gap states also give rise to two-step absorption (Fig. 44.1). We also cannot neglect photoinduced phenomena induced by these photoelectronic excitations (Sect. 44.4). Selecting an appropriate glass for a specific application is, therefore, not a straightforward process.

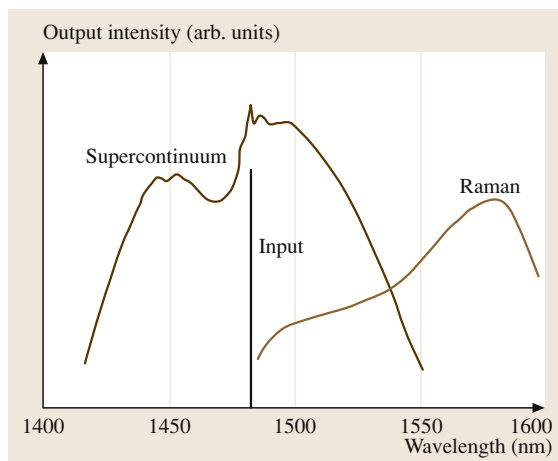


Fig. 44.6 Output spectra from a Raman fiber amplifier [44.38] and from a supercontinuum fiber generator [44.39] excited by 1483 nm light (schematic illustration)

44.1.3 Stimulated Light Scattering and Supercontinuum Generation

Recently, stimulated Raman scattering has attracted considerable interest, since it can be utilized for broadband light amplification [44.1, 5]. As illustrated in Fig. 44.1d, stimulated Raman scattering can be regarded as a kind of nondegenerate two-photon process. However, in contrast to two-photon absorption, one photon is absorbed by a virtual state at the same time as another photon is emitted in Raman scattering. As with conventional stimulated light emission, when the incident light is intense, the Stokes-shifted light can be amplified. The nonlinear polarization related to this is given by $P^{(3)}(\omega_s) \propto i\chi_R^{(3)} I(\omega_L) E(\omega_s)$, or the emitted light intensity is $\Delta I(\omega_s) \propto \chi_R^{(3)} I(\omega_L) I(\omega_s)$, where ω_L and $\omega_s (< \omega_L)$ are the exciting laser frequency and the Stokes-shifted frequency.

Most experiments have been performed using optical fibers [44.1, 42] in order to obtain long interaction lengths. For instance, as illustrated in Fig. 44.6, Raman fiber amplifiers comprising 5–10 km lengths of silica can provide broadband gains of $\approx 10 \text{ dB}$ [44.38]. Raman lasers using silica microspheres of diameter $\approx 70 \mu\text{m}$ [44.7] are also interesting examples. On the other hand, stimulated Brillouin scattering in glasses has also been explored [44.1, 4].

Supercontinuum generation in silica fibers has now been studied intensively [44.4, 39, 44]. When pulsed or CW light propagates through a nonlinear medium, it may undergo spectral broadening [44.5].

For instance, as illustrated in Fig. 44.6, a 350 m single-mode fiber excited by a 2.22 W CW laser with a wavelength of 1.483 μm can emit 2.1 W over a broad spectrum of 1.43–1.53 μm [44.39]. Note that, unlike the stimulated light scattering described above, light of a shorter wavelength is also generated in this process. Such a spectral-

conversion fiber could be utilized as a broadband optical amplifier. The phenomenon appears under strong and prolonged nonlinear interactions, and accordingly, several mechanisms such as intensity-dependent refractive index changes, third harmonic generation and stimulated Raman scattering could be responsible [44.39].

44.2 Second-Order Nonlinearity in Poled Glass

It has been discovered that several kinds of poling methods can add second-order nonlinearity to glasses [44.8]. At least five kinds have been demonstrated, which are listed in Table 44.2. Note that similar procedures are also employed for polymers. Most experiments utilize second harmonic signals to evaluate $\chi^{(2)}$ and, less commonly, electro-optical effects. Practical applications remain to be studied, while, for second-harmonic generations, the optical phase matching between exciting and nonlinearly-generated light is a prerequisite [44.45].

This so-called optical poling was demonstrated by *Österberg and Margulis* [44.9] using optical fibers. They found that exposing Ge-doped optical fibers of length ≈ 1 m to 70 kW Nd:YAG laser light for ≈ 1 h could increase the second-harmonic signal to 0.55 kW. *Stolen and Tom* [44.46] utilized two light beams (x1 and x2, of Nd:YAG laser light) for induction, which reduced the exposure time to ≈ 5 min. However, the method was only practical for optical fibers, since the nonlinearity induced was relatively small, $\chi^{(2)} \approx 10^{-4}$ pm/V. Second, so-called thermal poling, which was actually electrothermal in nature, was demonstrated in *bulk* SiO₂ samples by *Myers et al.* [44.47]. The nonlinearity induced, ≈ 1 pm/V, which was evaluated from the second harmonic signals of 1.06 μm laser light, is of a similar magnitude to that in quartz. Since this nonlinearity is reasonably large, this method has been widely applied to other glasses, such as PbO-SiO₂ [44.48], Nb₂O₅-B₂O₃-P₂O₅-CaO [44.49], and TeO₂-Bi₂O₃-ZnO [44.50]. Third, *Okada et al.* [44.51] demonstrated the corona-discharge poling at $\approx 200^\circ\text{C}$ of 7059 films deposited onto Pyrex glass substrates. This corona-poling procedure has often been employed for organic polymers. Fourth, electron-beam poling of PbO-silica glass was shown to produce ≈ 1 pm/V [44.52]. An advantage of this method is its high resolution, which may hold promise in the fabrication of optical integrated circuits, despite the fact that a vacuum is needed. *Liu et al.* (2001) [44.53] applied the method to chalcogenide glasses, which produced ≈ 1 pm/V. Proton implantation into silica can also add

Table 44.2 Reported poling methods, applied objects, typical procedures, and induced $\chi^{(2)}$ values in silica. For references, see the main text. The $\chi^{(2)}$ values listed are compared with $\chi_{11}^{(2)} = 1$ pm/V in crystalline SiO₂ and $\chi_{22}^{(2)} = 5$ pm/V in LiNbO₃

Method	Object	Procedure	$\chi^{(2)}$ (pm/V)
Optical	Fiber	Nd:YAG laser, 1 h	10^{-4}
Thermal	Bulk	4 kV, 300 $^\circ\text{C}$, 1 h	1
Corona	Film	5 kV, 300 $^\circ\text{C}$, 15 min	1
e-beam	Bulk	40 kV, 10 mA, 10 min	1
Proton	Bulk	500 kV, 1 mC, 100 s	1
UV	Bulk	ArF laser, 10 kV	3

a $\chi^{(2)}$ of ≈ 1 pm/V [44.54]. Sixth, *Fujiwara et al.* (1997) demonstrated UV poling in Ge-doped SiO₂ subjected to electric fields of $\approx 10^5$ V/cm. The $\chi^{(2)}$ induced is reported to be ≈ 3 pm/V, comparable to that in LiNbO₃.

Two poling mechanisms of note have been proposed [44.8]. One is that space charge produces a built-in electric field of $E_{\text{DC}} (\approx 10^6 \text{ V/cm})$, which induces an effective $\chi^{(2)}$ of $3E_{\text{DC}}\chi^{(3)}$ [44.46]. Here, E_{DC} is governed by the migration of ions such as Na⁺ under applied or generated electric fields [44.55]. In agreement with this idea, $\chi^{(2)}$ decays with a time constant of 10^1 – 10^6 days at room temperature, which is connected to the alkali ion mobility [44.55]. The other idea is that oriented defects such as E' centers are responsible. It is reasonable to assume that UV excitation produces defective dipoles, which are oriented with the static electric field.

It may be reasonable to assume that the dominant mechanism depends upon the poling method. Actually, we can divide the procedures listed in Table 44.2 into two types, depending upon whether or not the glass is heated during the poling process. The heating tends to enhance macroscopic ion migration,

while it can also relax microscopic defect orientations at the same time [44.25]. Therefore, it seems that the ion migration is responsible in thermal and coronal poling, while defect orientation dominates in the other methods. In this context, poling at low temperatures may be a promising way of enhancing defect orientation.

44.3 Particle-Embedded Systems

Glasses that contain nanoparticles of metals [44.65, 66] and semiconductors [44.67] have attracted considerable interest due to their unique third-order nonlinearities. Table 44.3 lists several recent results. Such glasses containing dispersed nanoparticles can be prepared by a variety of physical and chemical methods, e.g., vacuum deposition and sol-gel techniques [44.4, 65, 68]. For semiconductor systems, a lot of work has also been investigated in semiconductor-doped color glass filters [44.67], which are now commercially available.

These nanoparticle systems work efficiently at close to the resonant wavelengths of some electronic excitations. This feature produces at least three characteristics. First, the imaginary, not the real, part of $\chi^{(3)}$ may be more prominent. Accordingly, Table 44.3 compares absolute values. Second, the system exhibits a strong spectral dependence [44.4, 69]. For instance, Au-silica and CdSSe-silica are efficient at ≈ 580 nm and ≈ 800 nm [44.4]. Third, the response time τ and the linear absorption α tend to become longer and higher. Actually, a trade-off between $\chi^{(3)}$ and τ and α seems to exist. For instance, CdSSe-dispersed glasses show $\chi^{(3)}$ value of $\approx 10^{-9}$ esu with τ value of ≈ 20 ps, while Au-dispersed glasses give $\approx 10^{-11}$ esu and ≈ 1 ps, respectively [44.4]. Linear absorption can be as large as 10^4 cm $^{-1}$ [44.65], so these systems can be utilized as

In so-called *glass ceramics*, embedded crystals seem to be responsible for prominent $\chi^{(2)}$ [44.50, 56, 57]. For instance, *Takahashi* et al. [44.57] demonstrated that oriented Ba $_2$ TiSi $_2$ O $_8$ crystals are produced in BaO-TiO $_2$ -SiO $_2$ glass by heat treatment at 760 °C for 1 h, which gives a prominent $\chi^{(2)}$ of ≈ 10 pm/V. $\chi^{(2)}$ can also be generated at interfaces [44.58].

small devices, not as fibers. Note that, in pure silica at nonresonant infrared wavelengths, $\chi^{(3)} \approx 10^{-13}$ esu, $\tau \approx 10$ fs, and $\alpha \leq 10^{-5}$ cm $^{-1}$ [44.65]. As is suggested above, the particle-embedded system should surmount two problems for wide applications. One is the reduction of linear attenuation, including absorption and scattering, and the other is the shift of resonant wavelengths to the optical communication region, 1.3–1.5 μ m. Is such a shift possible? What are the mechanisms that give rise to these prominent nonlinearities in particle-embedded systems?

When the particle is a semiconductor such as CuCl and CdSSe, excitons or confined electron-hole pairs are responsible [44.67]. Specifically, the excitons in semiconductor particles behave as two-level systems, and at the resonance frequency, $|\chi^{(3)}|$ is written as

$$|\chi^{(3)}| \approx \text{Im}\chi^{(3)} \propto |\mu|^4 N T_1 T_2^2, \tag{44.9}$$

where μ is the dipole moment of the exciton, N is the particle number, T_1 (≈ 100 ps) is the lifetime of the exciton, and T_2 (\approx fs) is the dephasing time. A quantitative estimation predicts that closely packed CuCl particles of radius 40 nm could provide a $|\chi^{(3)}|$ enhancement of a factor of $\approx 10^3$ when compared with the bulk value [44.70].

When metal particles such as spherical Au particles with diameters of 10–50 nm are used, we can envisage

Table 44.3 Several recently reported particle systems, along with their preparation methods, $|\chi^{(3)}|$ values, and response times τ at the measured wavelength λ , as well as references. PLD and VE depict pulsed laser deposition and vacuum evaporation

System	Preparation	$ \chi^{(3)} $ (esu)	τ	λ (nm)	Reference
Au (15 nm)/ silica	Shell structure	10^{-9}	2 ps	550	[44.59]
Cu (2 nm)/Al $_2$ O $_3$	PLD	10^{-7}	5–450 ps	600	[44.60]
Ag (20 nm)/BaO	VE	10^{-10}	0.2 ps	820	[44.61]
Fe (4 nm)/BaTiO $_3$	PLD	10^{-6}		532	[44.62]
SnO $_2$ (10 nm)/silica	Sol-gel	10^{-12}		1064	[44.63]
CdS (4 nm)/silica	Sol-gel	10^{-11}		500	[44.64]

local field enhancement and dynamic responses from conduction electrons, including plasmon effects [44.65, 66, 71]. The result is approximately written as

$$\chi^{(3)} \approx p_m \chi_m^{(3)} |3\varepsilon_h/(\varepsilon_m + 2\varepsilon_h)|^2 \{3\varepsilon_h/(\varepsilon_m + 2\varepsilon_h)\}^2, \quad (44.10)$$

where p_m is the volume fraction of the metal particles, $\chi_m^{(3)}$ is the bulk nonlinearity of the metal, and ε_h and ε_m are the linear dielectric constants of the host (glass) and the metal. The nonlinearity of the host is neglected here for the sake of simplicity. Note that ε_h can be real, while ε_m is complex. We see that the metal's nonlinearity $\chi_m^{(3)}$ is decreased by the volume factor p_m , while $\chi_m^{(3)}$ may be enhanced by local fields if $\text{Re}(\varepsilon_m + 2\varepsilon_h) \approx 0$, which determines the resonance wavelength. In agreement with this model, an Au-dispersed glass, for instance, gives a greater $\chi^{(3)}$ than that of an Au film [44.5].

However, since these systems are complex, consisting of particles and a glass matrix, a variety of situations arise. For the particle, as well as particle species, we need to consider size, size distribution, shape and concentration. Here, needless to say, a narrow size

distribution is preferred for investigating fundamental mechanisms. The shape may be spherical, ellipsoidal, rod, and so forth. The concentration determines the mean separation distance between particles. As the concentration increases, electrical particle–particle interactions appear [44.59], and then the particles eventually percolate [44.60], which may produce dramatic changes in the nonlinear response. The matrix seems to be of secondary importance. Actually, liquids and organic materials have been employed as well as glasses [44.4]. However, some reports suggest that the nonlinearity is greatly affected by the surface states of the semiconductor particles [44.72] and by the dielectric constant of the matrix surrounding the metallic particles [44.4].

Lastly, it may be worth mentioning two recent results. One is that particles can be arrayed to produce photonic structures that exhibit novel nonlinear properties, such as light confinement [44.73]. The other is the generation of second harmonic signals from oriented ellipsoidal Ag nanoparticles [44.74]. Oriented particle structures have been produced in silica through tensile deformation and simultaneous heating, and this may be regarded as a kind of mechanical poling.

44.4 Photoinduced Phenomena

The intense pulsed light employed to produce nonlinear effects is also likely to produce a variety of transitory and (quasi-)stable optical changes [44.72, 75]. Photochromic effects induced by sub-gap light are the result of transitory electronic changes [44.76]. The increase in the refractive index of SiO₂ induced by fs–ns laser light

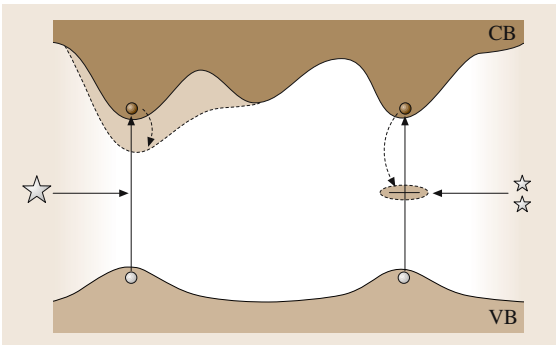


Fig. 44.7 Different relaxation paths from bandgap excitation due to one-photon (*left*) and two-photon (*right*) processes. For instance, the one-photon and two-photon excitations induce bonding strains and defect alterations

with $\hbar\omega \approx 1 \approx 5$ eV is known to occur along with quasi-stable structural changes [44.11, 77]. The optical poling described in Sect. 44.2 can be regarded as a kind of photoinduced anisotropy. When the light is more intense than 1 GW/cm^2 – 1 TW/cm^2 , depending on the duration of the light pulse, permanent damage is likely to occur [44.5]. These photosensitive changes are induced by electronic excitations, which may be triggered by one-photon and multiphoton processes, depending upon the photon energy and the light intensity.

Are these photoinduced phenomena really excited by nonlinear processes? For instance, when a photoinduced phenomenon is induced by exposure to light of photonic energy $\hbar\omega \approx E_g/m$, where m is an integer, it is often asserted that m -photon absorption is responsible. Nevertheless, the phenomenon may be triggered by one-photon absorption by mid-gap states, which give weak absorption tails (Fig. 44.2) arising from defective structures in the glass [44.25, 78]. Otherwise, it may be triggered by m -step absorption processes [44.79], consisting of m one-photon absorptions in succession (Fig. 44.1c). These processes have probably been overlooked when analyzing experimental results.

In addition, as schematically illustrated in Fig. 44.7, the structural changes induced by linear and nonlinear excitation are not necessarily the same. Tanaka has demonstrated using As_2S_3 that the bandgap excitations produced by one-photon and two-photon absorption produce different changes [44.78]. One-photon excitation leads to photodarkening, while two-photon excitation results in an increase in the refractive index with no photodarkening. In these phenomena, temperature rises upon intense exposure can be neglected. Such changes related to excitation processes can appear in glasses because the *localized* atomic wavefunction plays an important role, in contrast to the extended Bloch

wavefunctions in crystals. For instance, as we can see from (44.4), one-photon excitation occurs between wavefunctions with different parities, such as from s to p orbitals, since H is an odd function. In contrast, (44.5) shows that two-photon excitation can occur between states of the same parity states, such as from p to p. In addition, the $1/(E_{ni} - \hbar\omega)$ in the equation leads to the possibility of *resonant and localized* two-photon excitation if $E_{ni} - \hbar\omega \approx 0$ is satisfied for a mid-gap state. In such cases, the mid-gap state selectively absorbs the excitation energy, leading to an atomic change which shows itself as a macroscopic photoinduced phenomenon.

44.5 Summary

At present, there is a big gap between the science and the technology of glasses. From a physical point of view, glass science lags far behind crystal science, mainly due to experimental difficulties associated with atomic structure determination and the theoretical unavailability of Bloch-type wavefunctions. On the other hand, it is now difficult to imagine a world without optical fibers. These contrasting situations are promoting a deeper understanding of nonlinear photonic glasses and are leading to an increasingly wide range of applications for such glasses.

Third-order optical nonlinearity in homogeneous glasses has been studied in a fair amount of detail. Substantial experimental data have been obtained for n_2 , which have been analyzed using empirical relations such as those from Boling. In contrast, less work has been done on the nonlinear absorption β . In the present article, therefore, we have tried to present, in a coherent way, the absorption and the refractive index using semiconductor terminology. This approach can be used to connect nonlinear properties to the energy gap. However, little work has been done on the dynamics associated with this field. For instance, it is difficult to theoretically predict the response time τ in a glass at a particular excitation energy.

Two inhomogeneous systems have aroused increasing interest. One is poled or crystallized glass. Enhanced second-order nonlinearity has been reported for such systems, and this can sometimes be added selectively to a region. The other is the particle-embedded system, which can give greater third-order nonlinearity, and in some cases enhanced second-order nonlinearity too. However, the mechanism associated with this system remains to be studied. One experimental problem is to find a method that can reproduce particles of a fixed size and shape in arbitrary concentrations. These two inhomogeneous systems can be combined using photonic structure concepts, which will be of interest for future applications.

Finally, nonlinear optical excitations appear to play important roles in a variety of photoinduced phenomena in glasses. However, fundamental studies are still lacking in this area. The phenomenon of interest may be nonlinear in nature, or the linear excitation of gap states may trigger successive changes. Attempting to understand the nonlinear photo-electro-structural process will provide a challenging problem.

References

- | | | | |
|------|--|------|---|
| 44.1 | G. P. Agrawal: <i>Nonlinear Fiber Optics</i> , 3rd edn. (Academic, San Diego 2001) | 44.4 | R. L. Sutherland: <i>Handbook of Nonlinear Optics</i> , 2nd edn. (Marcel Dekker, New York 2003) |
| 44.2 | P. P. Mitra, J. B. Stark: <i>Nature</i> 441 , 1027 (2001) | 44.5 | R. W. Boyd: <i>Nonlinear Optics</i> , 2nd edn. (Academic, San Diego 2003) |
| 44.3 | L. F. Mollenauer: <i>Science</i> 302 , 996 (2003) | | |

- 44.6 M. H. Field, J. Popp, R. K. Chang: *Progress in Optics Vol. 41*, ed. by E. Wolf (North Holland, Amsterdam 2000) Chap. 1
- 44.7 S. M. Spillane, T. J. Kippenberg, K. J. Vahala: *Nature* **415**, 621 (2002)
- 44.8 Y. Quiquempois, P. Niay, M. Dounay, B. Pommellec: *Curr. Opin. Solid State Mater. Sci.* **7**, 89 (2003)
- 44.9 U. Österberg, W. Margulis: *Opt. Lett.* **11**, 516 (1986)
- 44.10 A. Jha, X. Liu, A. K. Kar, H. T. Bookey: *Curr. Opin. Solid State Mater. Sci.* **5**, 475 (2001)
- 44.11 G. Pacchioni, L. Skuja, D. L. Griscom: *Defects in SiO₂ and Related Dielectrics: Science and Technology* (Kluwer, Dordrecht 2000)
- 44.12 E. M. Vogel, M. J. Weber, D. M. Krol: *Phys. Chem. Glasses* **32**, 231 (1991)
- 44.13 K. Tanaka: *J. Mater. Sci. Mater. Electron.* **16**, 633 (2005)
- 44.14 M. E. Lines: *J. Appl. Phys.* **69**, 6876 (1991)
- 44.15 A. Zakery, S. R. Elliott: *J. Non-Cryst. Solids* **330**, 1 (2003)
- 44.16 K. Tanaka: *J. Non-Cryst. Solids* **338–340**, 534 (2004)
- 44.17 T. Mizunami, K. Takagi: *Opt. Lett.* **19**, 463 (1994)
- 44.18 A. Dragonmir, J. G. McInerney, N. Nikogosyan: *Appl. Opt.* **41**, 4365 (2002)
- 44.19 K. Tanaka, N. Yamada, M. Oto: *Appl. Phys. Lett.* **83**, 3012 (2003)
- 44.20 K. Tanaka: *Appl. Phys. Lett.* **80**, 177 (2002)
- 44.21 K. Ogusu, J. Yamasaki, S. Maeda, M. Kitao, M. Minakata: *Opt. Lett.* **29**, 265 (2004)
- 44.22 K. Imanishi, Y. Watanabe, T. Watanabe, T. Tsuchiya: *J. Non-Cryst. Solids* **259**, 139 (1999)
- 44.23 R. C. Enck: *Phys. Rev. Lett.* **31**, 220 (1973)
- 44.24 P. S. Weitzman, U. Osterberg: *J. Appl. Phys.* **79**, 8648 (1996)
- 44.25 S. R. Elliott: *Physics of Amorphous Materials*, 2nd edn. (Longman Scientific, Essex 1990)
- 44.26 N. Sugimoto, H. Kanbara, S. Fujiwara, K. Tanaka, K. Hirao: *Opt. Lett.* **21**, 1637 (1996)
- 44.27 S. Smolorz, I. Kang, F. Wise, B. G. Aiken, N. F. Borrelli: *J. Non-Cryst. Solids* **256 and 257**, 310 (1999)
- 44.28 O. Sugimoto, H. Nasu, J. Matsuoka, K. Kamiya: *J. Non-Cryst. Solids* **161**, 118 (1993)
- 44.29 N. L. Boling, A. J. Glass, A. Owyong: *IEEE Quant. Electron.* **14**, 601 (1978)
- 44.30 V. Dimitrov, S. Sakka: *J. Appl. Phys.* **79**, 1741 (1996)
- 44.31 J. Qi and D. F. Xue, G. L. Ning: *Phys. Chem. Glasses* **45**, 362 (2004)
- 44.32 S. Suehara, P. Thomas and A. Mirgorodsky, T. Merle-Mejean, J. C. Champarnaud-Mesjard, T. Aizawa, S. Hishita, S. Todoroki, T. Konishi, S. Inoue: *J. Non-Cryst. Solids* **345&346**, 730 (2004)
- 44.33 M. Sheik-Bahae, D. J. Hagan, E. W. Van Stryland: *Phys. Rev. Lett.* **65**, 96 (1990)
- 44.34 T. S. Moss: *Optical Properties of Semiconductors* (Butterworths, London 1959) p. 48
- 44.35 D. C. Hutchings, M. Sheik-Bahae, D. J. Hagan, E. W. Van Stryland: *Opt. Quantum. Electron.* **24**, 1 (1992)
- 44.36 M. Sheik-Bahae, D. C. Hutchings, D. J. Hagan, E. W. Van Stryland: *IEEE Quantum Electron.* **27**, 1296 (1991)
- 44.37 K. Tanaka: *Optoelectronic Materials and Devices*, Vol. 1, ed. by G. Lucovsky, M. Popescu (INOE, Bucharest 2004) Chap. 3
- 44.38 J. Bromage: *J. Lightwave Technol.* **22**, 79 (2004)
- 44.39 A. Zheltikov: *Appl. Phys. B* **77**, 143 (2003)
- 44.40 V. Mizrahi, K. W. DeLong, G. I. Stegeman, M. A. Saifi, M. J. Andrejco: *Opt. Lett.* **14**, 1140 (1989)
- 44.41 M. Asobe: *Opt. Fiber Technol.* **3**, 142 (1997)
- 44.42 G. I. Stegeman, R. H. Stolen: *J. Opt. Soc. Am. B* **6**, 652 (1989)
- 44.43 K. Tanaka, T. Gotoh, N. Yoshida, S. Nonomura: *J. Appl. Phys.* **91**, 125 (2002)
- 44.44 A. K. Abeeluck, C. Headley: *Appl. Phys. Lett.* **85**, 4863 (2004)
- 44.45 H.-Y. Chen, C.-L. Lin, Y.-H. Yang, S. Chao, H. Niu, C. T. Shih: *Appl. Phys. Lett.* **86**, 81107 (2005)
- 44.46 R. H. Stolen, H. W. K. Tom: *Opt. Lett.* **12**, 585 (1987)
- 44.47 R. A. Myers, N. Mukherjee, S. R. J. Brueck: *Opt. Lett.* **22**, 1732 (1991)
- 44.48 Y. Luo, A. Biswas, A. Frauenglass, S. R. Brueck: *Appl. Phys. Lett.* **84**, 4935 (2004)
- 44.49 B. Ferreira, E. Fargin, J. P. Manaud, G. Le Flem, V. Rodriguez, T. Buffeteau: *J. Non-Cryst. Solids* **343**, 121 (2004)
- 44.50 G. S. Murugan, T. Suzuki, Y. Ohishi, Y. Takahashi, Y. Benino, T. Fujiwara, T. Komatsu: *Appl. Phys. Lett.* **85**, 3405 (2004)
- 44.51 A. Okada, K. Ishii, K. Mito, K. Sasaki: *Appl. Phys. Lett.* **60**, 2853 (1992)
- 44.52 P. G. Kazansky, A. Kamal, P. St. Russell: *Opt. Lett.* **18**, 683 (1993)
- 44.53 Q. M. Liu, F. X. Gan, X. J. Zhao, K. Tanaka, A. Narazaki, K. Hirao: *Opt. Lett.* **26**, 1347 (2001)
- 44.54 L. J. Henry, B. V. McGrath, T. G. Alley, J. J. Kester: *J. Opt. Soc. Am. B* **13**, 827 (1996)
- 44.55 O. Deparis, C. Corbari, G. Kazansky, K. Sakaguchi: *Appl. Phys. Lett.* **84**, 4857 (2004)
- 44.56 V. Pruneri, P. G. Kazansky, D. Hewak, J. Wang, H. Takebe, D. N. Payne: *Appl. Phys. Lett.* **70**, 155 (1997)
- 44.57 Y. Takahashi, Y. Benino, T. Fujiwara, T. Komatsu: *Appl. Phys. Lett.* **81**, 223 (2002)
- 44.58 R. T. Hart, K. M. Ok, P. S. Halasyamani, J. W. Zwan-ziger: *Appl. Phys. Lett.* **85**, 938 (2004)
- 44.59 Y. Hamanaka, K. Fukuta, A. Nakamura, L. M. Liz-Marzan, P. Mulvaney: *Appl. Phys. Lett.* **84**, 4938 (2004)
- 44.60 R. Del Coso, J. Requejo-Isidro, J. Solis, J. Gonzalo, C. N. Afonso: *J. Appl. Phys.* **95**, 2755 (2004)
- 44.61 Q. F. Zhang, W. M. Liu, Z. Q. Xue, J. L. Wu, S. F. Wang, D. L. Wang, Q. H. Gong: *Appl. Phys. Lett.* **82**, 958 (2003)
- 44.62 W. Wang, G. Yang, Z. Chen, Y. Zhou, H. Lu, G. Yang: *J. Appl. Phys.* **92**, 7242 (2002)

44.63 A. Clementi, N. Chiodini, A. Paleari: Appl. Phys. Lett. **84**, 960 (2004)

44.64 S. G. Lu, Y. J. Yu, C. L. Mak, K. H. Wong, L. Y. Zhang, X. Yao: Microelectron Eng. **66**, 171 (2003)

44.65 F. Gonella, P. Mazzoldi: *Handbook of Nanostructured Materials and Nanotechnology*, Vol. 4, ed. by H. S. Nalwa (Academic, San Diego 2000) Chap. 2

44.66 V. M. Shalaev: *Nonlinear Optics of Random Media* (Springer, Berlin, Heidelberg 2000)

44.67 G. Banfi, V. Degiorgio, D. Ricard: Adv. Phys. **47**, 510 (1998)

44.68 M. Nogami, S. T. Selvan, H. Song: Photonic glasses: Nonlinear optical and spectral hole burning properties. In: *Handbook of Advanced Electronic and Photonic Materials and Devices*, Vol. 5, ed. by H. S. Nalwa (Academic, San Diego 2001) Chap. 5

44.69 J. He, W. Ji, G. H. Ma, S. H. Tang, H. I. Elim, W. X. Sun, Z. H. Zhang, W. S. Chin: J. Appl. Phys. **95**, 6381 (2004)

44.70 Y. Li, M. Takata, A. Nakamura: Phys. Rev. B **57**, 9193 (1998)

44.71 D. Stroud, P. M. Hui: Phys. Rev. B **37**, 8719 (1988)

44.72 A. Puzder, A. J. Williamson, F. Gygi, G. Galli: Phys. Rev. Lett. **92**, 217401 (2004)

44.73 M. Ajgaonkar, Y. Zhang, H. Grebel, C. W. White: Appl. Phys. Lett. **75**, 1532 (1999)

44.74 A. Podlipensky, J. Lange, G. Seifert, H. Graener, I. Cravetchi: Opt. Lett. **28**, 716 (2003)

44.75 R. C. Jin, Y. C. Cao, E. C. Hao, G. S. Metraux, G. C. Schatz, C. A. Mirkin: Nature **425**, 487 (2003)

44.76 Y. Watanabe, Y. Kikuchi, K. Imanishi, T. Tsuchiya: Mater. Sci. Eng. B **54**, 11 (1998)

44.77 J. S. Aitchison, J. D. Prohaska, E. M. Vogel: Met. Mater. Proc. **8**, 277 (1996)

44.78 K. Tanaka: Philos. Mag. Lett. **84**, 601 (2004)

44.79 K. Kajihara, Y. Ikuta, M. Hirano, H. Hosono: Appl. Phys. Lett. **81**, 3164 (2002)

45. Nonlinear Optoelectronic Materials

In a nonlinear optical material, intense light alters the real and imaginary components of the refractive index. The nonlinear response of the real part of refractive index modifies the phase of propagating light, while the imaginary part describes the change in absorption. These illumination-dependent properties of nonlinear materials provide the basis for all-optical switching—the ability to manipulate optical signals without the need for optical–electronic–optical conversion.

In this chapter we review the physical processes underlying the illumination-dependent refractive index. We review the real and imaginary nonlinear response of representative groups of materials: crystalline semiconductors, organic materials, and nanostructures, and we examine the practical applicability of these groups of materials to all-optical optical switching. We identify the spectral regions which offer the most favorable nonlinear response as characterized using engineering figures of merit.

45.1 Background	1075
45.1.1 Signal Processing in Optical Networks	1075
45.1.2 Optical Signal Processing Using Nonlinear Optics	1076

45.1.3 The Approach Taken During this Survey of Nonlinear Optoelectronic Materials	1076
--	------

45.2 Illumination-Dependent Refractive Index and Nonlinear Figures of Merit (FOM)	1077
45.2.1 Ultrafast Response	1077
45.2.2 Ultrafast Nonlinear Material Figures of Merit	1078
45.2.3 Resonant Response	1079
45.2.4 Resonant Nonlinear Material Figures of Merit	1079
45.3 Bulk and Multi-Quantum-Well (MQW) Inorganic Crystalline Semiconductors	1080
45.3.1 Resonant Nonlinearities	1080
45.3.2 Nonresonant Nonlinearities in Inorganic Crystalline Semiconductors	1083
45.4 Organic Materials	1084
45.4.1 Resonant Nonlinear Response of Organic Materials	1085
45.4.2 Nonresonant Nonlinear Response of Organic Materials	1086
45.5 Nanocrystals	1087
45.6 Other Nonlinear Materials	1088
45.7 Conclusions	1089
References	1089

45.1 Background

Optical fiber provides a suitable medium in which it is possible to reach tremendous transmission rates over long distances [45.1]. The maximum information-carrying capacity has been estimated to be around 100 Tbps [45.2]. Very high data rates can be achieved using a combination of wavelength- and time-division multiplexing techniques (WDM and TDM). WDM involves sending many signals in parallel at closely spaced wavelengths along the same fiber, while TDM allows close spacing in time of bits in a single channel.

While there exist means to produce, transfer, and detect information at a very high bandwidth, there is a need for more agility in photonic networks.

The agility of present-day optical networks is limited by the electronic nature of a very important function: the processing of information-bearing signals. Signal processing is responsible for switching and routing traffic, establishing links, restoring broken links, and monitoring and managing the network.

45.1.1 Signal Processing in Optical Networks

At present, the important and functionally complex signal-processing operations of switching and routing are carried out electronically. Electronic signal processing imposes two significant limitations on the functionality of optical networks: cost and opacity. Signal switching and routing requires conversion of the optical information into electrical signals, processing in the electronic domain, and converting back to the optical domain before retransmission. Such an operation requires detection, retiming, reshaping, and regeneration at each switching and routing point. This necessitates complex and expensive electronic and electro-optical hardware at each routing and switching node. The use of electronic signal processing places strict requirements on the format of data streams transferred and processed, thus making the signal processing opaque. Repetition rates of optical signals, power levels, and packet lengths have to be standardized before they can be processed electronically.

The ability to perform signal processing operations entirely within the optical domain would eliminate the requirement for optical–electrical–optical conversions, while providing agility and speed inherent to optical elements. Optical signal processing, in contrast with electronics, may provide ultrafast sub-picosecond switching times [45.3].

45.1.2 Optical Signal Processing Using Nonlinear Optics

Nonlinear optics can potentially support ultrafast self-processing of signals.

A variety of nonlinear optical signal-processing functions can be realized with similar fundamental building blocks [45.4–6]. Nonlinear optical elements and devices can be either integrated in photonic circuits [45.7] or used in a free-standing configuration [45.8]. Nonlinear optics can enable signal processing without the requirement for external electrical, mechanical, or thermal control [45.9]. The response time of properly designed nonlinear optical devices is limited fundamentally only by the nonlinear response time of the constituent materials [45.3, 10–12].

Photons do not interact with each other in vacuo. In order to perform nonlinear optical signal-processing operations the properties of a medium through which the light travels must be modified by the light itself. Optical signals then propagate differently as a result of their influence on the medium. Nonlinear optical signal-processing elements utilize the illumination-dependent

real and imaginary parts of the index of refraction [45.9]. Depending on the material and spectral position, the real part of the refractive index and absorption of a given nonlinear material can either increase or decrease with increasing illumination.

A wide range of broadband and wavelength-selective nonlinear optical signal-processing devices has been proposed and demonstrated.

The most commonly studied nonlinear optical switching elements are nonlinear Fabry–Perot interferometers, nonlinear Mach–Zehnder modulators, nonlinear directional couplers, optical limiters, and nonlinear periodic structures.

A nonlinear Fabry–Perot interferometer consists of two mirrors separated by a nonlinear material. As the refractive index of the nonlinear material changes with an increased level of illumination, the effective path length of the resonator is altered. A nonlinear Fabry–Perot interferometer can be tuned out of, or into, its transmission resonance. When illuminated with the continuous-wave light, a nonlinear Fabry–Perot interferometer can exhibit optical bistability. Optical bistability is a phenomenon in which the instantaneous transmittance of the device depends both on the level of incident illumination and on the prior transmittance of the device. Such an element enables all-optical memory.

In a nonlinear Mach–Zehnder modulator and a nonlinear directional coupler, a part of the waveguide is made out of a nonlinear material. Changing the intensity of the incident light changes the effective path length experienced by the light. This, in turn, through phase interference, results in an illumination-dependent transmittance in a Mach–Zehnder modulator, and an illumination-dependent coupling in a nonlinear directional coupler.

A number of techniques use nonlinear properties of materials to obtain power-limiting, and associated with it, on–off switching. Such devices are based on total internal reflection [45.13], self-focusing [45.14], self-defocusing, two-photon absorption [45.15], or photorefractive beam fanning [45.16].

Nonlinear periodic structures combine the phenomena of nonlinear index change and distributed Bragg reflection. The intensity-dependent transmission and reflection properties of nonlinear periodic structures can be harnessed to yield various signal-processing functions. It has been demonstrated that nonlinear periodic structures can support optical switching and limiting [45.4–6, 17–20], optical bistability [45.8, 21–25], solitonic propagation of pulses [45.26, 27], and pulse compression [45.28].

45.1.3 The Approach Taken During this Survey of Nonlinear Optoelectronic Materials

There exist excellent texts that describe nonlinear optical processes and review the published properties of nonlinear materials [45.29–35]. This chapter will discuss the applicability of different material groups to non-

linear optical switching. Following the introduction of the concept of nonlinear refractive index and figures of merit in Sect. 45.2, the nonlinear properties of inorganic crystalline semiconductors (Sect. 45.3), organic materials (Sect. 45.4), nanocrystals (Sect. 45.5), and selected other materials (Sect. 45.6) will be reviewed and summarized. A critical review is given with a focus on figures of merit and processability.

45.2 Illumination-Dependent Refractive Index and Nonlinear Figures of Merit (FOM)

In a nonlinear optical medium intense light alters the real and imaginary components of the refractive index. The nonlinear response of the real part of the refractive index modifies the phase of propagating light, while the imaginary part describes the change in absorption.

This subsection will present the formalism used to describe how light affects the ultrafast and resonant changes in the nonlinear refractive index. The ultrafast nonlinear index changes take place in the spectral region where the material is nonabsorbing, while the resonant nonlinear index changes take place in the absorbing spectral region.

45.2.1 Ultrafast Response

Ultrafast nonlinear response is characterized by the instantaneous response, weak nonlinear index changes, and weak nonlinear absorption. The formalism that describes the ultrafast changes in the real and imaginary parts of the refractive index can be derived from the theory of nonlinear polarization.

The polarization $\mathbf{P}(r, \omega)$ of a material in the presence of an electric field $\mathbf{E}(r, \omega)$ at a frequency ω and position r is defined as

$$\mathbf{P}(r, \omega) = \epsilon_0 \chi(r, \omega) \mathbf{E}(r, \omega), \quad (45.1)$$

where ϵ_0 is the permittivity of free space and $\chi(r, \omega)$ is the dielectric susceptibility tensor. $\chi(r, \omega)$ is related to the index of refraction $n(\omega)$ by

$$\chi(r, \omega) = n^2(r, \omega) - 1. \quad (45.2)$$

In a homogeneous nonlinear material $\chi(r, \omega) = \chi(\omega)$ but $\chi(\omega)$ is not constant with electric field and the influence of $\mathbf{E}(r, \omega)$ on $\mathbf{P}(r, \omega)$ is not linear. In this case it is customary to expand $\mathbf{P}(r, \omega)$ in a power series of

$\mathbf{E}(r, \omega)$ [45.34]:

$$\begin{aligned} \mathbf{P}(r, \omega) = & \epsilon_0 \chi^{(1)}(\omega) \mathbf{E}(\omega) \\ & + \epsilon_0 \left[D^{(2)} \sum_{j,k} \chi_{ijk}^{(2)}(-\omega_3; \omega_1, \omega_2) \right. \\ & \times \mathbf{E}_j(\omega_1) \mathbf{E}_k(\omega_2) \\ & + D^{(3)} \sum_{jkl} \chi_{ijkl}^{(3)}(-\omega_4; \omega_1, \omega_2, \omega_3) \\ & \times \mathbf{E}_j(\omega_1) \mathbf{E}_k(\omega_2) \mathbf{E}_l(\omega_3) \\ & \left. + \text{higher-order terms} \right], \end{aligned} \quad (45.3)$$

where $\chi^{(1)}$ is the linear susceptibility, while $\chi^{(2)}$ and $\chi^{(3)}$ are the coefficients of the second- and third-order nonlinear susceptibility. The coefficients $D^{(2)}$ and $D^{(3)}$ are defined as:

$$D^{(2)} = \begin{cases} 1, & \text{for indistinguishable fields} \\ 2, & \text{for distinguishable fields} \end{cases} \quad (45.4)$$

and

$$D^{(3)} = \begin{cases} 1, & \text{for all fields indistinguishable} \\ 2, & \text{for two fields indistinguishable} \\ 3, & \text{for all fields distinguishable} \end{cases} \quad (45.5)$$

In all known materials the higher-order components of the effective nonlinear susceptibility tensor $\chi(\omega)$ yield smaller contributions to the effective polarization than the preceding terms of the same parity. On the other hand, in the presence of strong electric field the terms designated as the *higher-order terms* in (45.3) [i. e. terms proportional to the powers of $\mathbf{E}(r, \omega)$ higher than four],

can be larger than the first three terms. However, the assumption of moderate intensities and the aim to illustrate the concept of nonlinear refractive index justifies retaining only the first three terms of (45.3) in the derivation that follows.

Nonlinear optical switching relies on nonlinear effects in which intense light changes the refractive index. Under such conditions there are no direct-current (DC) or low-frequency electro-optic effects present and the second term in (45.3) can be neglected. In addition all values of ω are degenerate. $\mathbf{P}(r, \omega)$ reduces to

$$\begin{aligned}\mathbf{P}(r, \omega) &= \epsilon_0 [\chi^{(1)}(\omega) + \chi^{(3)}(\omega) \mathbf{E}(\omega) \mathbf{E}(\omega)] \mathbf{E}(\omega) \\ &= \epsilon_0 \left[\chi^{(1)}(\omega) + \frac{2\chi^{(3)}(\omega) I}{\epsilon_0 n_0 c} \right] \mathbf{E}(\omega),\end{aligned}\quad (45.6)$$

where I is the local intensity

$$I = \frac{\epsilon_0}{2} n_0 c |E(\omega)|^2, \quad (45.7)$$

and c is the speed of light in vacuum.

The first term in (45.6) represents the linear contribution to the polarization and the second term represents the nonlinear, intensity-dependent part. This intensity-dependent part gives rise to the nonlinear index of refraction fundamental to this work.

To obtain the direct expression for the nonlinear refractive index the effective susceptibility from (45.6) is substituted into (45.2) to yield

$$n^2 = 1 + \chi^{(1)} + \frac{2\chi^{(3)}(\omega) I}{\epsilon_0 n_0 c}. \quad (45.8)$$

In order to relate directly this nonlinear part of polarization to the intensity-dependent part of the refractive index – a macroscopic measurable quantity – the effective index of refraction is expressed as

$$n = n_0 + n_2 I. \quad (45.9)$$

Taking the square of (45.9) and neglecting the terms proportional to I^2 under the assumption of weak relative nonlinearity ($n_2^2 I^2 \ll n_0 n_2 I$ and $n_2^2 I^2 \ll n_0^2$) gives

$$n^2 = n_0^2 + 2n_0 n_2 I. \quad (45.10)$$

Equating (45.8) and (45.10) gives an expression for n_2

$$n_2 = \frac{\chi^{(3)}}{\epsilon_0 n_0^2 c}, \quad (45.11)$$

where all the factors are in SI units.

In general, n_2 can have real (Re) and imaginary (Im) parts with $n_{2\text{Re}}$ responsible for the nonlinear refraction

and $n_{2\text{Im}}$ responsible for the nonlinear absorption or gain. There are many conventions used to express the real and imaginary parts of the nonlinear refractive index. The approach used by researchers must always be determined prior to comparison with absolute numbers. However, in general it is safe to write

$$n_{2\text{Re}} = \frac{K}{n_0^2} \text{Re}(\chi^{(3)}) \quad (45.12)$$

and

$$n_{2\text{Im}} = \frac{K}{n_0^2} \text{Im}(\chi^{(3)}), \quad (45.13)$$

where the constant K depends on the convention and units used [45.34].

In the rest of this chapter n_2 will be used to express the real part of the ultrafast nonlinear index of refraction, i.e. n_2 will be as used in (45.12).

In order to account for the imaginary component of the ultrafast nonlinear response in a commonly used way the following relationship is defined

$$\alpha = \alpha_0 + \beta I. \quad (45.14)$$

Equation (45.14) expresses the total absorption (α) in terms of its linear (α_0) and nonlinear (βI) contributions. β is the measurable, macroscopic quantity that will be used throughout this chapter to quantify the effects of the ultrafast imaginary nonlinear response.

45.2.2 Ultrafast Nonlinear Material Figures of Merit

A nonlinear material useful in a nonlinear optical signal-processing device must simultaneously satisfy the following conditions:

- The excitation time of the nonlinear effect must be less than the pulse width.
- The sum of the excitation and the relaxation times must be shorter than the pulse spacing.

In addition, an ultrafast nonlinear material must satisfy the following requirements:

- The effect of linear absorption must be weak compared to the effect of nonlinear refraction. *Stegeman* quantifies this condition in terms of the unitless figure of merit W [45.12]

$$W = \frac{|\Delta n|}{\alpha_0 \lambda} > 1, \quad (45.15)$$

where Δn is the induced change in the real part of the refractive index, α_0 is the linear absorption (expressed in units of inverse length) and λ

is the wavelength of light (with units of length). To facilitate consistent comparison between different nonlinear materials, Δn in (45.15) was assumed to be evaluated as the intensity approaches the saturation intensity, at which the rate of change of the refractive index drops noticeably below a linear dependence on intensity [45.12]. In general (45.15) can be used to quantify the nonlinear quality of a given material at any intensity, not only at the saturation.

- The effect of two-photon absorption must be weak compared to the effect of nonlinear refraction. This condition is quantified using the figure of merit T [45.12]

$$T = \frac{\beta_2 \lambda}{n_2} < 1, \quad (45.16)$$

where β is the two-photon absorption coefficient from (45.14) (expressed in units of length/power).

The conditions (45.15) and (45.16) can be combined in terms of a single figure of merit F

$$F = \frac{|\Delta n|}{\alpha_{\text{eff}} \lambda} > 1, \quad (45.17)$$

where α_{eff} is the effective absorption experienced by the sample at a given intensity. F can be used to quantify the quality of materials for signal processing with respect to nonlinear processes of any order rather than with respect to only third-order processes as in (45.16).

Condition (45.17) ensures that the nonlinear phase shift $\Delta\phi^{\text{NL}} = 2\pi\Delta nL/\lambda$, where L is the length of the material, reaches 2π before the intensity decays to $1/e$ of its input value as a result of the effective absorption. Phase shifts between 0.5π and 3.5π are required for most optical switching devices [45.12].

45.2.3 Resonant Response

The resonant response of a nonlinear material is the dominant nonlinear effect in the linearly absorbing spectral region. A different formalism than that presented in Sects. 45.2.1 and 45.2.2 is used to describe the resonant changes in the real and imaginary parts of the refractive index.

Illumination with light which is resonant with the material results in the direct absorption of the incoming photons, generating excited states and giving rise to a decrease in the effective absorption. If the relaxation time of the excited states is longer than the length of the pulse, the resonant effect is proportional to the

fluence, rather than to the intensity of the incident ultrafast pulse. This saturation of the absorption is described by the following expression for the effective absorption α_{eff} [45.35]

$$\alpha_{\text{eff}} = \frac{\alpha'_0}{1 + \frac{P}{P_{\text{sat}}}} + \alpha_u. \quad (45.18)$$

where α_u is the unsaturable absorption, $P = \int_0^t I(t') dt'$ is the incident fluence and P_{sat} is the saturation fluence at which the effective absorption decreases to half of its initial value $\alpha_0 = \alpha'_0 + \alpha_u$. P accounts for the cumulative (up to the duration of the pulse) character of the resonant nonlinear response.

The saturation of absorption is accompanied by a change in the real part of the refractive index [45.35]

$$\Delta n = \frac{n'_2 P}{1 + \frac{P}{P_{\text{sat}}}}. \quad (45.19)$$

n'_2 describes the strength of the real part of the resonant nonlinear refractive index.

In this chapter, nonresonant and resonant phenomena are considered. The parameters n_2 and β from Sect. 45.2.1 are used to quantify the ultrafast response and Δn and $\Delta\alpha$ from Sect. 45.2.1 are used to describe the resonant response.

45.2.4 Resonant Nonlinear Material Figures of Merit

Figures of merit for the nonresonant response have been defined in Sect. 45.2.2. This section will introduce resonant figures of merit that account for the nonlinear phase shift that accumulates over the duration of a pulse.

For illustrative purposes, first-order approximations to (45.18) and (45.19) of the form $\Delta n(t) = \int_0^t n'_2 I(t') dt'$ and $\alpha_{\text{eff}}(t) = \alpha_0$ are considered under the assumption $P \ll P_{\text{sat}}$. A resonant nonlinear material is assumed to be illuminated with a square pulse of the form:

$$I(t) = \begin{cases} I_0, & \text{if } 0 < t < \tau_p \\ 0, & \text{if } t > \tau_p. \end{cases} \quad (45.20)$$

In analogy to (45.17) a time-averaged nonlinear figure of merit is defined for the resonant response

$$\langle F \rangle = \frac{|\langle \Delta n \rangle|}{\langle \alpha \rangle \lambda}. \quad (45.21)$$

The time-averaged nonlinear index change is

$$\begin{aligned}\langle \Delta n \rangle &= \frac{1}{\tau_p} \int_0^{\tau_p} \left[\int_0^t n'_2 I(t') dt' \right] dt \\ &= \frac{n'_2 P_{\text{total}}}{2} = \frac{|\Delta n_{\text{peak}}^{\text{ultrafast}}|}{2},\end{aligned}\quad (45.22)$$

where P_{total} is the total fluence of the pulse $P_{\text{total}} =$

$\int_0^{\tau_p} I(t') dt' = I_0 \tau_p$. The time-averaged absorption is $\langle \alpha \rangle = \alpha_0$. For the case considered, the figure of merit (45.21) becomes:

$$\langle F \rangle = \frac{|\Delta n_{\text{peak}}|}{2\alpha_0 \lambda}, \quad (45.23)$$

which is half of the ultrafast figure of merit. For simplicity, (45.17) will be used throughout this chapter for both resonant and ultrafast response.

45.3 Bulk and Multi-Quantum-Well (MQW) Inorganic Crystalline Semiconductors

The illumination-dependent refractive and absorptive nonlinear properties of inorganic crystalline semiconductors have been studied comprehensively. Since semiconductors are at the heart of the electronics industry, semiconductor micro- and nanofabrication techniques are well established. This enables the preparation of high-quality nonlinear samples and devices. The ability to change the composition of semiconductor compounds allows the tuning of the electronic band gap over the visible and infrared spectral ranges. The spectral position of the band gap, in turn, tunes the nonlinear properties.

When a semiconductor is illuminated with light at a frequency within the absorbing region, the dominant nonlinear effect relies on the presence of linear absorption. Upon absorption of the incident light, the electrons undergo a transition from the valence band to the conduction band, saturating the absorption. This band-filling effect is accompanied by a very large change in the real part of the refractive index.

In a nonresonant nonlinear process no single-photon absorption takes place. Under illumination with intense light the electronic clouds of the constituent atoms are distorted, changing the refractive index of the material. Associated with this is a multiphoton absorption process which takes place when the sum of the photon energies is larger than the band-gap energy. This effect changes the absorption characteristics of the material. Both the real and imaginary parts of ultrafast nonlinear index change, given their connection through the nonlinear Kramers–Kronig relations.

In addition, when subjected to an intense continuous wave or a high-repetition-rate pulsed illumination, the temperature of the absorbing materials, including semiconductors, increases. This in turn changes the refractive index. Thermal effects have relaxation times as long as

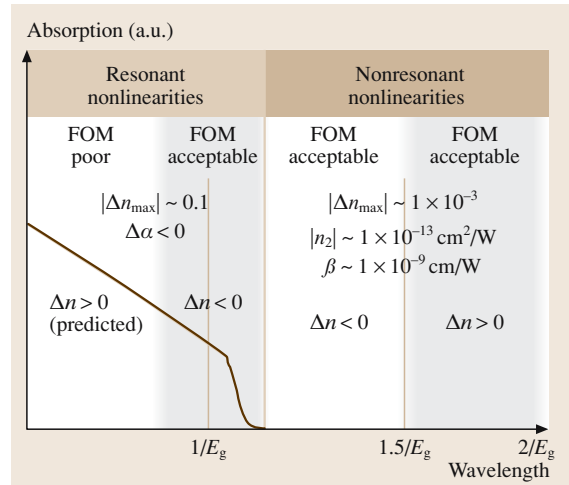


Fig. 45.1 Trends in the nonlinear response of bulk semiconductors

milliseconds and are not useful in processing trains of closely spaced pulses.

Figure 45.1 shows typical trends in the nonthermal nonlinear response of bulk inorganic crystalline semiconductor material under picosecond, low-repetition-rate illumination.

45.3.1 Resonant Nonlinearities

The two most important characteristics of the resonant nonlinear response are saturation of absorption and large nonlinear index change [45.36–38]. The relaxation of resonant nonlinear effects in semiconductors is not instantaneous. As long as the duration of the incident pulse is shorter than the relaxation time of the material, the magnitude of a nonlinear resonant response

is proportional to the fluence, rather than to the intensity of the incident pulse. The relatively long relaxation time of nonlinear effects in bulk and multi-quantum-well (MQW) inorganic crystalline semiconductors (from hundreds of picoseconds to tens of nanoseconds) is often used as an argument against using resonant nonlinearities. However, established techniques such as low-temperature growth and doping can reduce the relaxation time down to tens of picosecond [45.39, 40].

The phenomenon of saturation of absorption ($\Delta\alpha < 0$) translates into absorption that decreases with increasing incident fluence. Resonant figures of merit of semiconductors are acceptable near the band edge and become worse at lower wavelengths. Such behavior is due to a stronger, lower-threshold saturation of absorption around the band edge.

Nonlinear index change is negative around the band edge and has been predicted to be positive at wavelengths lower than those corresponding to the first heavy-hole and light-hole excitonic peaks.

Resonant Nonlinearities in Bulk Semiconductors

In 1991 Gupta et al. measured the time of nonlinear response of GaAs grown at low temperatures (LT-GaAs) [45.42]. Changes in reflectivity were monitored during a pump-probe experiment at 620 nm. Relaxation times of several 2- μm -thick samples grown at temperatures ranging between 190 °C and 400 °C were measured. A decrease in the decay time to 0.4 ps was recorded with decreasing growth temperatures. This short relaxation time is drastically lower than the typical value of nanoseconds for unannealed GaAs [45.42].

In 1993 Harmon et al. studied the dependence of the nonlinear relaxation time in LT-GaAs on annealing temperatures. A decrease in the relaxation time down to sub-picosecond values was observed with decreasing annealing temperatures [45.43].

In a number of papers published between 1994 and 1998, the group of Smith, Othonos, Benjamin, and Loka reported on a series of comprehensive experiments carried out on various LT-GaAs samples grown using molecular beam epitaxy (MBE). The dependence of the magnitude and the response time of nonlinear effects on the growth and annealing temperatures was studied. Very large negative nonlinear index changes were measured ($\Delta n_{\text{max}} = -0.13$) accompanied by a strong saturation of absorption [45.44]. The relaxation time was measured to decrease to a few picoseconds for samples grown at 500 °C [45.45]. The pump-probe measurements were carried out in the band-edge region at

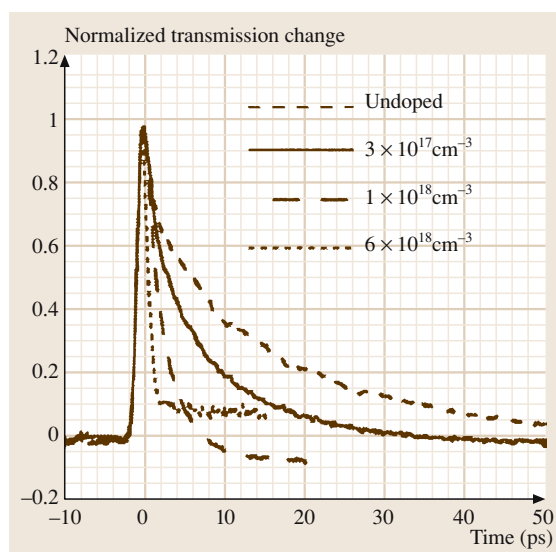


Fig. 45.2 Results of the pump-probe measurements illustrating the time-resolved change in transmission for He-InGaAsP samples with different Be doping concentrations. (After [45.41] with permission)

wavelengths of 870–890 nm. The decreased response time was attributed to the fast decay of excited carriers to mid-gap states. These states are an effect of the LT growth [45.40].

In subsequent years another group of researchers studied the strength and dynamics of intensity-dependent response in InGaAsP doped with Be grown with He-plasma-assisted MBE. As illustrated in Fig. 45.2, sub-picosecond relaxation times were obtained [45.41]. The rapid decay was explained by a short lifetime of excited states due to the existence of mid-gap He and He-Be trap states. Again, large negative changes in the real part of the refractive index and strong saturation of absorption were observed [45.39].

Resonant Nonlinearities in Semiconductor Multi-Quantum-Wells

The nonlinear properties of semiconductor multi-quantum-wells (MQWs) are similar to those of bulk semiconductors [45.38, 46, 47]. The nonlinear response in MQWs around the band edge is stronger and begins at lower fluences than in bulk materials.

Compared with bulk semiconductors, semiconductor MQWs offer an additional degree of freedom in selecting their nonlinear properties. The effective electronic band gap of a given semiconductor MQW structure, and hence the dispersion of real and imaginary

parts of its linear and nonlinear refractive index, are influenced by two factors: the choice of the compositions of the constituent compounds and the well-to-barrier thickness ratio.

In 1982 Miller et al. reported on the measurements of resonant nonlinear properties of semiconductor MQWs. A very strong absorption saturation was noticed around the first excitonic peak in GaAs/AlGaAs MQWs. Based on these results a large refractive nonlinearity was deduced from the nonlinear Kramers–Kronig relation [45.48]. A theoretical paper followed, explaining the dynamics of transient excitonic nonlinearities [45.49]. A 20-ns excited-carrier relaxation time was predicted.

In 1986 Lee et al. measured the nonlinear saturation of the absorption of bulk GaAs and 29.9-nm GaAs/AlGaAs wells grown by molecular beam epitaxy. The measurement was performed using a monochromatic pump and a broadband probe over a 40-nm spectral range near the MQW band edge. Using the nonlinear Kramers–Kronig relation, large index changes of both signs were predicted. In MQWs, absorptive and refractive nonlinearities were enhanced compared to bulk GaAs. Index changes ranging from $\Delta n = -0.06$ to $\Delta n = 0.03$ were predicted in the samples analyzed [45.50].

This report was followed in 1988 by a study of nonlinearities around the band edge carried out by the

same research group [45.38]. The response of bulk GaAs was compared with that of three sets of GaAs/AlGaAs MQWs, with well thicknesses of 7.6 nm, 15.2 nm and 29.9 nm. Again, a strong saturation of absorption was measured and nonlinear index changes of both signs were predicted from the nonlinear Kramers–Kronig relation [45.51]. The magnitude of the change in the real part of the refractive index was predicted to increase with decreasing well size. The sign of the refractive nonlinearity changed at wavelengths slightly shorter than that corresponding to the first excitonic peak [45.38].

Since 1988 many results of research on the nonlinear properties of GaAs/AlGaAs MQWs have been reported by Garmire et al. In a series of papers, the saturation of absorption was studied in GaAs/AlGaAs MQWs grown by metalorganic chemical vapor deposition epitaxy. The nonlinear Kramers–Kronig relation was used to predict the associated change in the real part of the refractive index. Figure 45.3 shows the predicted enhancement of nonlinearity with decreasing well size and the change of sign near the excitonic peak. Attempts were made to use the illumination-dependent shift of Fabry–Perot fringes to estimate directly the negative nonlinear index change along the band edge. However, this approach was admitted to yield significant errors, with the Fabry–Perot technique sometimes giving a value of Δn at twice the magnitude predicted

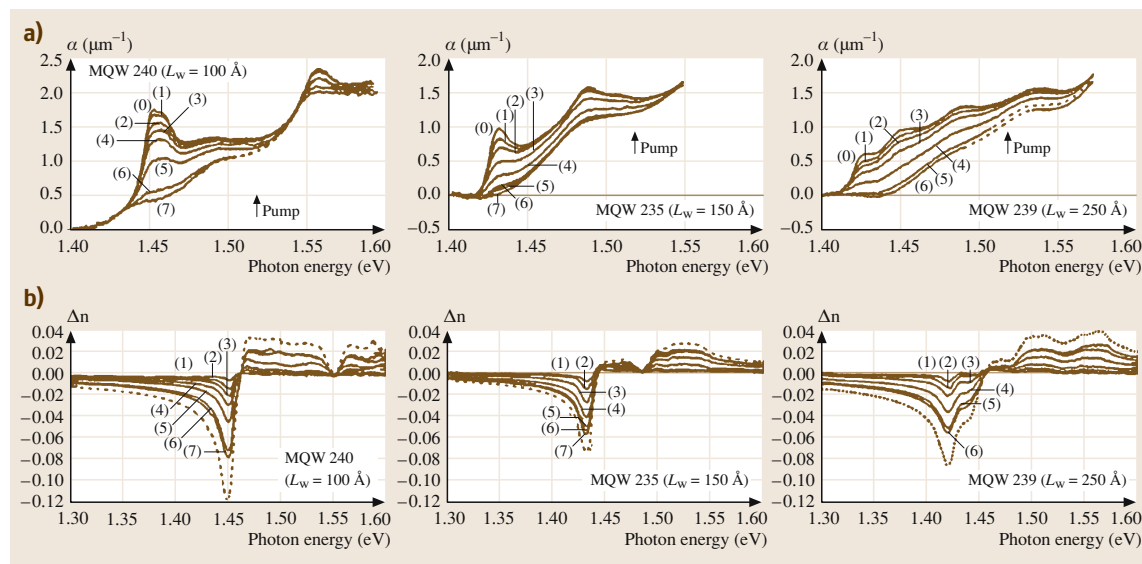


Fig. 45.3a,b Spectra of GaAs/AlGaAs MQWs of three different well widths, measured at various incident intensities by pulsed pump at 1.52 eV: **(a)** absorption coefficient α ; **(b)** the change in the real part of the refractive index Δn . (After [45.46] with permission)

from the Kramers–Kronig relation [45.46]. In 1987 Fox et al. reported nonlinear measurements around the band edge of bulk GaInAs [45.52] and GaInAs/InP MQWs near wavelengths of 1.6 μm [45.53]. Full saturation of absorption was observed. The nonlinear index changes deduced from the nonlinear Kramers–Kronig relation were slightly larger than that observed in GaAs [45.35].

Recently Brzozowski et al. published results of direct picosecond measurements of nonlinear refractive-index change and nonlinear absorption in $\text{In}_{0.530}\text{Al}_{0.141}\text{Ga}_{0.329}\text{As}/\text{In}_{0.530}\text{Ga}_{0.470}\text{As}$ multi-quantum-wells in the range 1480–1550 nm. Large low-threshold nonlinear index changes were found: Δn of up to 0.14 with a figure of merit of 1.38 at a fluence of 116 $\mu\text{J}/\text{cm}^2$. The figure of merit F was greater than unity over much of the spectrum. The results are summarized in Fig. 45.4.

In 1996 Judawlikis et al. reported the decreased nonlinear relaxation time in LT-grown Be-doped InGaAs/InAlAs MQWs. Nonlinear relaxation times of a few tens of picoseconds were observed in a pump-probe experiment near the band edge. The nonlinear change in the real part of the refractive index was not reported [45.56].

A different approach to decrease the response times of band-edge nonlinearities of semiconductor MQWs was taken by the groups of White, Sibbet, and Adams. An electric current was applied to active InGaAsP/InP waveguides and the nonlinear optical response under electrical bias was studied. It was found that under

a forward bias the refractive nonlinear response was quenched. Under a reverse bias the nonlinear response was slightly reduced, but the initially long recovery time was reduced to 50 ps [45.57] and 18 ps in subsequent experiments [45.58]. Further, it was found that, when the waveguide was biased at transparency, the nonlinear coefficients of the semiconductor MQW waveguides were $n_2 = 4 \times 10^{-11} \text{ cm}^2/\text{W}$ and $\beta = 4 \times 10^{-9} \text{ cm}/\text{W}$, giving a combined figure of merit of $F = 7$ [45.59]. In all measurements the negative nonlinear index changes were measured to have magnitude smaller than $|\Delta n| < 0.001$ [45.60].

45.3.2 Nonresonant Nonlinearities in Inorganic Crystalline Semiconductors

Nonresonant nonlinearities are not triggered by direct electronic transitions due to single photons. Much weaker effects of distortion of electronic clouds and multi-photon absorption are responsible for nonresonant nonlinear response. Maximum nonresonant nonlinear index changes are of the order $|\Delta n_{\text{max}}| \approx 1 \times 10^{-3}$. Since in certain spectral regions a typical nonresonant Kerr coefficient is $n_2 \approx 1 \times 10^{-13} \text{ cm}^2/\text{W}$, the linear absorption is around 5 cm^{-1} , and the corresponding two-photon absorption coefficient is $\beta \approx 1 \times 10^{-9} \text{ cm}/\text{W}$, the figures of merit associated with nonresonant semiconductor nonlinearities can be acceptable.

The biggest advantage of nonresonant semiconductor nonlinearities is their sub-picosecond response time. The sum of the rise and relaxation times of nonresonant nonlinearity has been argued to be comparable to the orbital period of an electron in its motion about the nucleus, estimated to be around 10^{-16} s [45.29].

Depending on the spectral region, bulk and MQW inorganic crystalline semiconductors may exhibit either positive or negative refractive nonresonant nonlinearities. Under illumination with sub-nanosecond pulses at low repetition rates, the nonlinear index change is negative for wavelengths up to $1.5ch/E_g$, where ch/E_g is the wavelength corresponding to the band gap, and h is Planck's constant. Δn is positive for wavelengths longer than $1.5ch/E_g$ [45.37, 61]. In MQWs the spectral position of the sign change in Δn depends on the nanostructure of MQWs [45.62]. In 1993 Shaw and Jaros predicted through theory the dispersion of refractive nonlinearity in semiconductor MQWs and superlattices. They found that in MQWs the proximity of the spectral position of the Δn sign change to the band edge increases with increasing quantum confinement [45.62].

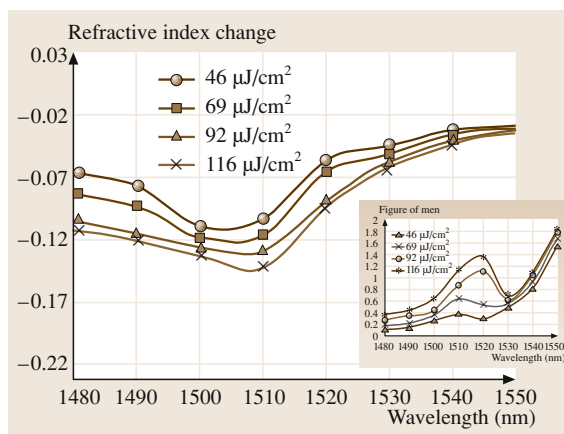


Fig. 45.4 Nonlinear index changes of 121 10-nm $\text{In}_{0.530}\text{Al}_{0.141}\text{Ga}_{0.329}\text{As}$ barriers and 120 5-nm $\text{In}_{0.530}\text{Ga}_{0.470}\text{As}$ wells grown on S-doped 001 InP wafer substrate. The inset shows the corresponding figures of merit. (After [45.54, 55])

Under nonresonant illumination with pulses longer than one nanosecond, there is no sign change in the refractive nonlinearity. The negative nonlinearity originating from two-photon absorption-induced free-carrier effects is much stronger than any positive third-order refractive effects at moderate and high intensities for $hc/E_g < \lambda < 2hc/E_g$. Consequently, the measured Δn is always negative in this spectral range [45.63].

The group of Sheik-Bahae and Van Stryland has authored several reports on the prediction of the spectral dependence of nonresonant nonlinearities in semiconductors. In 1985 Van Stryland et al. predicted trends in the absorptive ultrafast nonlinear response of semiconductors. An equation for the two-photon absorption below the band gap was derived and compared with experimental values. Dispersion of two-photon absorption is expected to mimic the dispersion of linear absorption; i. e. two-photon absorption is strong and relatively flat from the band gap to almost the midpoint of the band gap, at which point it goes to zero. Good agreement was obtained between experiment and theory for photon energies not in the vicinity of the band gap, with two-photon absorption coefficients of various semiconductors ranging from $\beta = 3 \times 10^{-9}$ cm/W to $\beta = 25 \times 10^{-9}$ cm/W [45.36].

In the ensuing years the same research group reported theory describing the spectral dependence of the real part of the ultrafast nonlinearity and compared it with experiments. The results are shown in Fig. 45.5. The magnitude of n_2 is largest near the photon energy corresponding to half of the band gap. Since, for wavelengths longer than that corresponding to half the band gap, two-

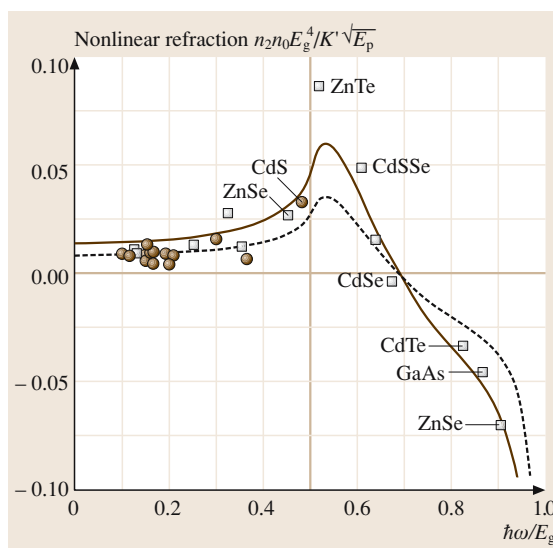


Fig. 45.5 Dispersion of the refractive nonlinearity of inorganic crystalline semiconductors in the transparent region. The points correspond to the experimental data explained in [45.37], while the lines are a fit to the theory. (After [45.37])

photon absorption vanishes, large figures of merit can be expected in these spectral region. In addition, n_2 was predicted to be positive for wavelengths longer than that corresponding to $0.75E_g$, and negative between $0.75E_g$ and E_g [45.37, 64]. A large discrepancy between theoretical and experimental results was observed near the band gap, where the theory drastically underestimated the strength of the refractive nonlinearity.

45.4 Organic Materials

Organic materials constitute another class of promising nonlinear materials. Organic materials exhibit significant nonlinearities across the visible and infrared spectral regions [45.65]. They are readily processable into thin-film waveguide structures [45.33, 66] and in general do not rely on a high degree of perfection in ordering or purity to manifest their desired properties. The molecules that make up organic materials provide a tremendous range of structural, conformational, and orientational degrees of freedom for exploration with the aid of novel synthetic chemistry. This permits flexible modification and optimization of linear and nonlinear properties [45.33].

As is the case with semiconductor nonlinearities, the nonlinear response of organic materials can be divided into resonant and nonresonant parts, occurring in the absorbing and transparent regions, respectively. The resonant nonlinearities are a result of a single-photon absorption, while the nonresonant nonlinearities arise as a result of perturbations of electronic clouds and multi-photon absorption.

Depending on the structure of the constituent molecules, organic materials may exhibit many absorption resonances and hence many spectral areas of different strength and sign of nonlinear response. Phenomena such as molecular reorientation and pho-

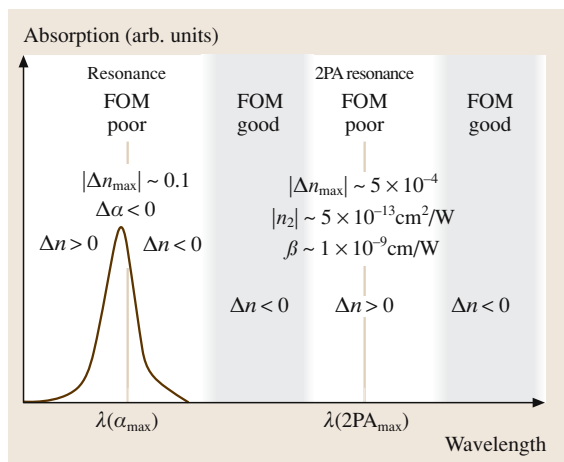


Fig. 45.6 Typical trends in the nonlinear response of organic materials with one absorption resonance

toisomerization, which are often found in organic materials, make the picture even more complex.

Some organic materials, such as most nonlinear dyes, have only one absorption resonance, which permits the qualitative prediction of their nonlinear response in the visible and near-infrared regions. Fig. 45.6 shows the nonlinear response of a typical nonlinear organic material with one absorption resonance.

In general, the figures of merit of organic materials in the absorbing region are poor. However, molecular effects such as *trans*–*cis* photoisomerization strongly increase the nonlinear index change along the absorption edge. The magnitudes of ultrafast nonlinearities and associated figures of merit of organic materials are comparable to those of inorganic crystalline semiconductors.

45.4.1 Resonant Nonlinear Response of Organic Materials

Since 1998 *Rangel and Rojo* et al. have reported resonant nonlinear properties of various organic compounds. The nonlinear refraction and absorption effects were reported in and near the absorbing regions of: solid-state PMMA samples doped with nonlinear azobenzene dye Disperse Red 1 [45.67], polydiacetylene microcrystals in aqueous suspension [45.68], cyclohexane suspensions of vanadium-oxide phthalocyanine microcrystals [45.69, 70], solid-state samples of polythiophene/selenophene copolymer [45.71], and a chloroform solution of triazole-quinone derivation [45.72].

PMMA films doped with Disperse Red 1 (10% molar concentration) have been shown to exhibit large,

low-threshold nonlinear index changes and saturation of absorption as a result of optically induced structural changes in the middle and near the edge of the absorption resonance at 490 nm [45.67]. The nonlinear index changes associated with this photochemical phenomenon, called *trans*–*cis* photoisomerization have exceeded 0.1 in the spectral region $\lambda < 590$ nm. As illustrated in Fig. 45.7 [45.67], light near the main absorption resonance causes the azobenzene molecule to change from the *trans* to the *cis* configuration. During this process, the distance between the two carbons from which the acceptor and donor groups extend reduces from about 9.0 Å to 5.5 Å. This results in a drastic reduction in the molecule's dipole moment, which reduces the material's polarizability, providing a large negative nonlinearity with nonlinear index changes reaching $\Delta n_{\text{max}} = 0.12$ at 560 nm under illumination with 20-ps pulses at a repetition rate of 10 Hz [45.67]. The figures of merit calculated according to (45.17) did not exceed 0.42 over the spectral range studied.

The refractive and absorptive nonlinear properties of polydiacetylene microcrystals in aqueous solution have been analyzed across their absorbing and near resonance regions 500–800 nm [45.68] with the same system as in [45.67]. As shown in Fig. 45.8, both negative and pos-

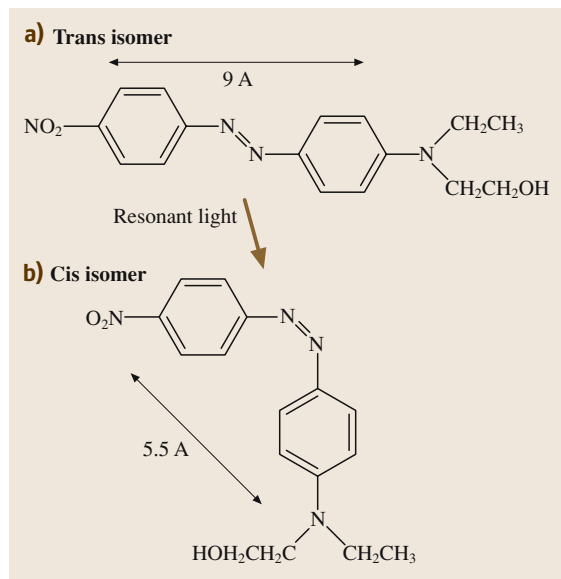


Fig. 45.7 Chemical structure of the azobenzene dye Disperse Red 1 undergoing *trans*–*cis* photoisomerization. Following resonant absorption, the azobenzene molecule changes its configuration, resulting in a decreased dipole moment. (After [45.67] with permission)

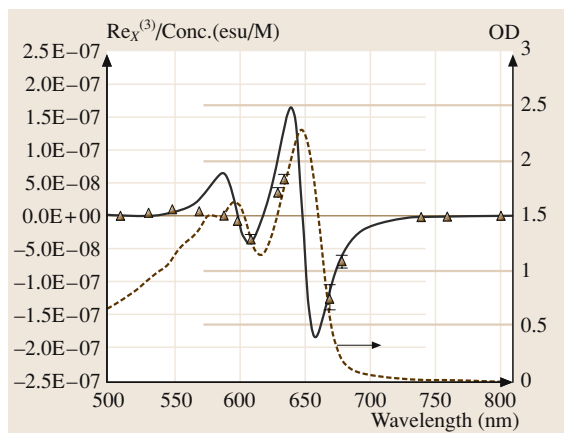


Fig. 45.8 $\chi^{(3)}$ and linear absorbance as a function of wavelength for 100-nm polydiacetylene microcrystals. (After [45.68])

itive nonlinear index changes were observed, depending on the spectral position relative to one of the two absorption resonances in the visible region. Saturation of absorption was observed across the entire absorbing region. A three-level model was developed to explain the data. Because of the low concentration of microcrystals in water, the absolute values of the nonlinear index change did not exceed 2×10^{-4} , while the figures of merit stayed below 0.26.

Cyclohexane suspensions of vanadium-oxide phthalocyanine microcrystals have also been characterized around its main absorption resonance peak in the spectral range 600–680 nm [45.69, 70]. Similarly to the nonlinear response of polydiacetylene microcrystals, both signs of refractive nonlinearity were observed, with the sign of the nonlinearity tracing the derivative of the linear absorption. Solid-state samples of polythiophene/selenophene copolymer were analyzed near the absorbing region and have shown only negative nonlinearity [45.71]. These results contradicted the predictions from nonlinear Kramers–Kronig transformation, which predicts positive nonlinearity in the spectral region where absorption increases with wavelength. The origins of this discrepancy were not understood.

In 1997 Demenicis et al. reported on the measurements of nonlinear properties of poly(3-hexadecylthiophene) in a chloroform solution around the absorption edge of 532 nm with 70-ps pulses at a repetition rate of 5 Hz [45.73]. Saturation of absorption and negative nonlinearity was observed. The saturation intensity decreased with increasing concentration, while the nonlinear absorption and nonlinear refraction at low

intensity increased linearly. Figures of merit were not given.

45.4.2 Nonresonant Nonlinear Response of Organic Materials

In the 1970s it was predicted from theory and experimentally demonstrated that the conjugation of organic molecules results in a strong electronic delocalization and an associated large third-order nonlinearity in the transparent spectral region. This region extends between the main absorption peak due to the band-to-band transition and the first of the vibronic modes of the conjugated chain [45.74, 75]. It was demonstrated that the solid-state polymerization of organic monomers results in $\chi^{(3)}$ values comparable to these observed in inorganic crystalline semiconductors. Numerous studies followed that concentrated on the determination of the length and bond order of the conjugation chain, and acceptor and donor strengths that yield the strongest nonlinear response. Marder et al. provide an excellent summary of the field of nonresonant nonlinear response of organic molecules in [45.76].

Single-crystal polydiacetylene para-toluene sulfonates (PTS) have received special attention among the organic nonlinear compounds. In 1994 Lawrence et al. reported that at 1600 nm PTS has a large nonlinear refractive index of $n_2 = 2.2 \times 10^{-12} \text{ cm}^2/\text{W}$ [45.77]. Since linear absorption of PTS is very low at this wavelength, and two-photon absorption was below experimental sensitivity ($\beta < 0.5 \text{ cm}/\text{GW}$), and very large figures of merit were predicted with W exceeding unity for incident intensities as low as $20 \text{ MW}/\text{cm}^2$, and T never exceeding 0.1. A report followed in which Lawrence et al. showed measured two-photon absorption and nonlinear refraction values of PTS in the spectral range 800–1600 nm [45.78]. The two-photon-absorption coefficient varied in the range $0\text{--}700 \text{ cm}/\text{GW}$ while the measured n_2 coefficient was between $n_2 = -2.2 \times 10^{-12} \text{ cm}^2/\text{W}$ and $n_2 = 4.3 \times 10^{-12} \text{ cm}^2/\text{W}$. In 2000 and 2003 Yoshino et al. published two reports where the influence of three- and four-photon absorption on the nonlinear response was studied in the near-infrared region [45.79, 80]. For wavelengths of 1600–2200 nm the nonlinear refraction coefficient n_2 was around $n_2 = 5 \times 10^{-13} \text{ cm}^2/\text{W}$ while nonlinear absorption was dominated by three- and four-photon effects.

Although PTS exhibits nonlinear refractive-index changes and figures of merit that are very large for non-

linear material in the transparent region, **PTS** suffers from low processability – as a single-crystal it cannot be easily processed into desired shapes.

Third-order nonresonant nonlinear properties of fullerene organic compounds have also attracted significant attention [45.81, 82]. Recently *Chen et al.*

reported nonlinear coefficients and figures of merit of high-quality polyurethane films heavily loaded with (60)fullerene (C_{60}) [45.83]. Nonlinear refractive coefficients in excess of $10^{-12} \text{ cm}^2/\text{W}$ were reported in the wavelength range 1150–1600 nm with very good figures of merit.

45.5 Nanocrystals

Nanoscale quantum-confined inorganic crystalline semiconductors represent an interesting group of nonlinear materials [45.84]. The size of such quantum dots is less than the bulk radii of excitons, holes, and electrons in a given semiconductor. As in the case of semiconductor **MQWs**, this results in quantum confinement of carriers. In a nanocrystal, this takes place in all three dimensions [45.85, 86]. Quantized energy levels make nanocrystals an artificial analogue of noninteracting atoms in a gas, raising the possibility of explaining the nonlinear processes by adopting the models of atomic physics.

To allow processability nanocrystals are usually embedded in either solid or liquid, organic or glass, optically linear hosts. Nanocrystal material systems are thus hybrids of semiconducting and insulating materials and combine interesting properties from both material groups. As in the case of semiconductor **MQWs**, the composition and size of quantum dots determines the energy of the electronic transitions. This allows spectral

tunability of absorption features and nonlinear properties over the entire visible and infrared spectrum. On the other hand, the organic or glass host permits flexible fabrication of samples, waveguides, and other integrated components using polymer photonics technologies [45.87].

Figure 45.9 shows the properties of a typical resonant and nonresonant nonlinear response of strongly confined semiconductor nanocrystal composites. The data presented in this figure are based on the published theoretical predictions and experimental reports.

The finite number of allowed lower electronic levels leads to more pronounced excitonic features and resonant nonlinearities, which take place at lower fluences than in bulk or **MQW** inorganic crystalline semiconductors [45.88].

Similarly to the nonlinear response of bulk and **MQW** semiconductors, the resonant nonlinear response of nanocrystals is characterized by the saturation of absorption and an associated large change in the real part of the refractive index.

Saturation of absorption in strongly confined PbS quantum-dot glasses was measured in the spectral range 1.2–1.3 μm [45.89], covering the spectral position of the valley between the first and second excitonic peaks in the 6.6-nm-diameter sample studied. This material system was used as a passive saturable absorber in the production of 4.6-ps pulses via mode-locking around a wavelength of 1.3 μm [45.89]. The report was followed by studies of saturation of absorption dynamics in quantum dots of various sizes at a wavelength of 1.3 μm . This wavelength covered spectral regions ranging from the first to second electronic transitions, depending on the size of nanocrystals. The saturation energy and nonlinear decay times at a given wavelength were found to decrease with increasing size of nanocrystals [45.90]. Values for the refractive nonlinearity were not reported.

Saturation of absorption in PbS quantum-dot-doped glasses was also studied under illumination with 70-ps and 15-ns pulses 1.06 μm [45.91]. The

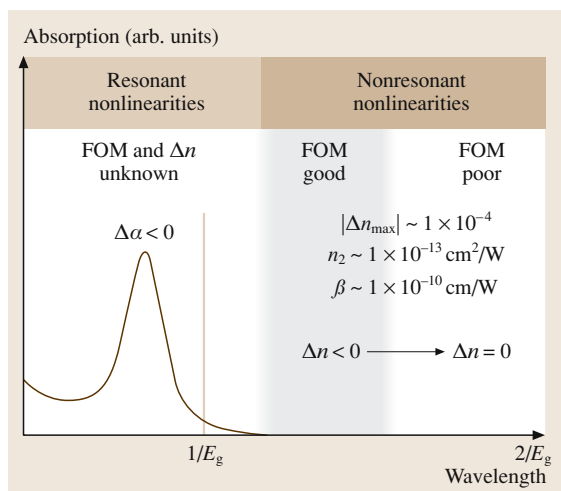


Fig. 45.9 Trends in the picosecond nonlinear response of inorganic semiconductor nanocrystals

nanocrystals analyzed had mean radii of 1.7–2.7 nm, resulting in excitonic peaks at wavelengths in the range 1.0–1.5 μm . The saturation intensity of the for the samples analyzed was found to be 2.3 MW/cm² and the relaxation time was measured to be 23 ± 2 ps.

Lu et al. measured the resonant nonlinear refractive properties of strongly confined 3.3-nm-diameter PbS nanocrystals in polymeric coatings over the spectral range 580–630 nm. This spectral range corresponds to the wavelengths around the first excitonic peak. The degenerate four-wave-mixing technique was used to measure the nonlinear susceptibility of nanocrystals near the photon energy of the first electronic transition at 595 nm. The values were found to fall in the range $\chi^{(3)} = 1 \times 10^{-6} - 1 \times 10^{-5}$ esu. Combined with the 50 kW/cm² intensity used in the experiment these values of $\chi^{(3)}$ suggest maximum nonlinear index changes of $\Delta n \approx 0.013$. No data on saturation of the absorption were reported.

The nonresonant nonlinear response of nanocrystals shows different dispersion characteristics than that of any other material group. Under illumination with picosecond pulses, the nonresonant third-order nonlinearity is negative for photon energies between half the band gap and the band gap [45.92]. The magnitude of

the third-order nonlinear response increases with proximity to the band gap, and disappears entirely near the half-band-gap energy. Such a response is in contrast to the Δn sign change between absorption and two-photon absorption resonances as observed in bulk semiconductors and organic materials. However, similarly to the nonresonant response of bulk semiconductors, the sign of the nonresonant refractive nonlinearity in the semiconductor-doped glasses in some spectral regions depends on the duration of the pulses used to measure the nonlinear effect. For pulses shorter than 1 ps the contribution of third-order positive refractive effects is comparable to that of the negative free-carrier absorption nonlinearities of the fifth order. Consequently, the measured Δn can be positive [45.63].

In 2000 and 2001 Liu et al. published several reports on measurements of the ultrafast nonlinearities of PbS nanoparticles, and PbS-coated CdS nanocomposites [45.93–95]. Surprisingly large refractive nonlinear indices of refraction up to -5×10^{-12} cm²/W and unmeasurable nonlinear absorption were observed in surface modified polymer–PbS composites at moderate concentrations of 1.9×10^{-3} mol/l. This large refractive nonlinearities were attributed to the surface recombination owing to the high surface-to-volume ratio of PbS nanoparticles.

45.6 Other Nonlinear Materials

Metallic nanocomposites and cascaded second-order materials are two other promising groups of nonlinear materials. Metalorganic nanocomposites are made out of metallic quantum dots embedded in organic or glass hosts. Resonant nonlinear properties of copper [45.96] and silver nanoparticles [45.97, 98] embedded in a glass host were measured using degenerate four-wave-mixing experiments at visible wavelengths. The nonlinear coefficients and figures of merit of the metallic nanocrystals characterized were similar to those of semiconductor nanocrystals and reached a maximum near the plasma-frequency absorption peak.

It has been argued that metallic nanoparticles can potentially exhibit stronger nonlinear effects than other

material systems [45.99]. This is associated with local field effects that enhance nonlinear response of the composite systems if the refractive index of the nonlinear constituent is lower than that of the linear host. Such a scenario can be realized in metallic nanoparticle–glass composites, since around the spectral positions of the plasma resonance the refractive index of metals can be lower than 1.

Cascaded nonlinear materials are made out of materials with second-order nonlinear properties. An appropriate design of such structures design results in a net accumulated phase shift for the illumination at a fundamental optical frequency at the end of a cascaded system. Cascaded material system acts as an effective third-order nonlinear material [45.100–102].

45.7 Conclusions

Following the preceding review, this section will summarize the major conclusions, as well as the missing pieces, of the published literature on nonlinear optical materials.

Bulk and MQW semiconductors have been demonstrated to exhibit low-threshold saturation of absorption near the band edge. The spectral position of the band edge can be tuned over the entire visible and near-infrared spectrum. It has been predicted from the nonlinear Kramers–Kronig relation, and has been measured directly in isolated cases, that the band-edge saturation of absorption results in large changes of the real part of refractive index.

In the regions of transparency, semiconductors exhibit weak nonlinear refractive effects of both signs. The nonresonant effects can be accompanied by two-photon absorption. Depending on the spectral position, the nonresonant nonlinear response of semiconductors can be characterized by good figures of merit.

Semiconductor nanocrystals also permit spectral tunability of their linear and nonlinear optical properties over the entire visible and near-infrared regions. Semiconductor nanocrystals have been demonstrated to exhibit strong saturation of absorption near the excitonic peak associated with the first allowed elec-

tronic transition. The nonresonant nonlinear response of nanocrystals is of a similar magnitude as in bulk and in MQW inorganic crystalline semiconductors.

The figures of merit for organic materials in the absorbing region are in general poor. In the transparent region the Kerr and two-photon absorption coefficients of organic materials are of magnitudes comparable to these of inorganic crystalline semiconductors. The sign of the refractive nonlinearity varies across the spectrum depending on the proximity to various absorption resonances.

Although the nonlinear properties of many materials systems have been reported, further characterization is needed to assess the applicability of various nonlinear material systems to optical signal processing. In contrast to previously reported measurements carried out at isolated wavelengths, measurements of the refractive and absorptive nonlinear response over wide spectral ranges, which would permit determination of figures of merit, need to be carried out. In particular, the refractive and absorptive nonlinear response in the most-promising absorption-edge regions of MQW semiconductors and semiconductor nanocrystals should be examined comprehensively and the applicability of these material systems to optical signal processing should be determined.

References

- 45.1 E. Cotter, J. K. Lucek, D. D. Marcenac: *IEEE Commun. Mag.* **34**, 90–95 (1997)
- 45.2 P. P. Mitra, J. B. Stark: *Nature* **411**, 1027 (2001)
- 45.3 P. W. Smith: *The Bell Syst. Tech. J.* **61**, 1975–1983 (1982)
- 45.4 L. Brzozowski, E. H. Sargent: *J. Opt. Soc. Am. B* **17**, 1360–1365 (2000)
- 45.5 L. Brzozowski, E. H. Sargent: *IEEE J. Quantum Electron.* **36**, 550–555 (2000)
- 45.6 L. Brzozowski, E. H. Sargent: *IEEE J. Quantum Electron.* **36**, 1237–1242 (2000)
- 45.7 P. W. Smith, I. P. Kaminov, P. J. Maloney, L. W. Stulz: *Appl. Phys. Lett.* **34**, 62–65 (1979)
- 45.8 P. W. Smith, E. H. Turner: *Appl. Phys. Lett.* **30**, 280–281 (1977)
- 45.9 B. E. A. Saleh, M. C. Teich: *Fundamentals of Photonics* (Wiley, New York 1991)
- 45.10 P. W. E. Smith, L. Qian: *IEEE Circuits Dev. Mag.* **15**, 28–33 (1999)
- 45.11 P. W. E. Smith: All-optical devices: materials requirements. In: *Nonlinear Optical Properties of Advanced Materials*, Vol. 1852 (SPIE, Los Angeles, CA 1993) pp. 2–9
- 45.12 G. I. Stegeman: All-optical devices: materials requirements. In: *Nonlinear Optical Properties of Advanced Materials*, Vol. 1852 (SPIE, Los Angeles, CA 1993) pp. 75–89
- 45.13 I. C. Khoo, M. Wood, B. D. Guenther: *Nonlinear liquid crystal optical fiber array for all-optical switching/limiting*, In: *LEOS '96 9th Annual Meeting*, Vol. 2, pp. 211–212, IEEE, (Bellingham, 1996)
- 45.14 G. L. Wood, W. W. II. I. Clark, M. J. Miller, G. J. Salamo, E. J. Sharp: Evaluation of passive optical limiters and switches. In: *Materials for Optical Switches, Isolators, and Limiters*, Vol. 1105 (SPIE, Orlando, FL 1989) pp. 154–181
- 45.15 R. Bozio, M. Meneghetti, R. Signorini, M. Maggini, G. Scorrano, M. Prato, G. Brusatin, M. Guglielmi: Optical limiting of fullerene derivatives embedded in sol-gel materials, In: *Photoactive Organic Materials. Science and Applications*, Proc. NATO Adv.

- Res. Workshop, Avignon, France, June 25–30, 1995 (Kluwer, Dordrecht 1996) 159–174
- 45.16 J. A. Hermann, P. B. Chapple, J. Staromlynska, P. Wilson: Design criteria for optical power limiters. In: *Nonlinear Optical Materials for Switching and Limiting*, Vol. 2229, ed. by M. J. Soileau (SPIE, Orlando, FL 1994) pp. 167–178
- 45.17 N. G. R. Broderick, D. Taverner, D. J. Richardson: Opt. Express **3**, 447–453 (1998)
- 45.18 N. D. Sankey, D. F. Prelewitz, T. G. Brown: Appl. Phys. Lett. **60**, 1427–1429 (1992.)
- 45.19 L. Brzozowski, E. H. Sargent: IEEE J. Lightwave Technol. **19**, 114–119 (2000)
- 45.20 L. Brzozowski, V. Sukhovatkin, E. H. Sargent, A. SpringThorpe, M. Extavour: IEEE J. Quantum Electron. **39**, 924–930 (2003)
- 45.21 H. M. Gibbs, S. L. McCall, T. N. C. Venkatesan, A. C. Gossard, A. Passner, W. Wiegmann: Appl. Phys. Lett. **35**, 451–453 (1979)
- 45.22 H. M. Gibbs, S. S. Tang, J. L. Jewell, D. A. Winberger, K. Tai, A. C. Gossard, S. L. McCall, A. Passner: Appl. Phys. Lett. **41**, 221–222 (1982)
- 45.23 H. G. Winful, J. H. Marburger, E. Garmire: Appl. Phys. Lett. **35**, 379–381 (1979)
- 45.24 D. Pelinovsky, L. Brzozowski, E. H. Sargent: Phys. Rev. E **60**, R4536–R4539 (2000)
- 45.25 D. Pelinovsky, J. Sears, L. Brzozowski, E. H. Sargent: J. Opt. Soc. Am. B **19**, 45–53 (2002)
- 45.26 C. M. de Sterke, J. E. Sipe: Progress Opt. **33**, 203–260 (1994)
- 45.27 W. Chen, D. L. Mills: Phys. Rev. Lett. **58**, 160–163 (1987)
- 45.28 W. N. Ye, L. Brzozowski, E. H. Sargent, D. Pelinovsky: J. Opt. Soc. Am. B **20**, 695–705 (2003)
- 45.29 R. W. Boyd: *Nonlinear Optics* (Academic, New York 1992)
- 45.30 M. G. Kuzyk, C. W. Dirk: *Characterization Techniques and Tabulations for Organic Nonlinear Optical Materials* (Dekker, New York, N.Y. 1998)
- 45.31 D. L. Mills: *Nonlinear Optics: Basic Concepts* (Springer, Berlin, Heidelberg 1998)
- 45.32 P. Gunter: *Nonlinear Optical Effects and Materials* (Springer, Berlin, Heidelberg 2000)
- 45.33 P. N. Prasad, D. J. Williams: *Introduction to Nonlinear Optical Effects in Molecules and Polymers* (Wiley, New York 1991)
- 45.34 R. L. Sutherland: *Handbook of Nonlinear Optics* (Dekker, New York 1996)
- 45.35 E. Garmire: IEEE J. Selected Topics Quantum Electron. **6**, 1094–1110 (2000)
- 45.36 E. W. Van Stryland, M. A. Woodall, H. Vanherzeele, M. J. Soileau: Opt. Lett. **10**, 490–492 (1985)
- 45.37 M. Sheik-Bahae, D. C. Hutchings, D. J. Hagan, E. W. Van Stryland: IEEE J. Quantum Electron. **27**, 1296–1309 (1991)
- 45.38 S. H. Park, J. F. Morhange, A. D. Jeffery, R. A. Morgan, A. Chavez-Pirson, H. M. Gibbs, S. W. Koch, N. Peyghambarian, M. Derstine, A. C. Gossard, J. H. English, W. Wiegmann: Appl. Phys. Lett. **52**, 1201–1203 (1988)
- 45.39 L. Qian, S. D. Benjamin, P. W. E. Smith, H. Pinkney, B. J. Robinson, D. A. Thompson: Opt. Lett. **22**, 108–110 (1997)
- 45.40 H. S. Loka, S. D. Benjamin, P. W. E. Smith: IEEE J. Quantum Electron. **34**, 1426–1437 (1998)
- 45.41 H. Pinkney, D. A. Thompson, B. J. Robinson, L. Qian, S. D. Benjamin, P. W. E. Smith: J. Cryst. Growth **209**, 237–241 (2000)
- 45.42 S. Gupta, M. Y. Frankel, J. A. Valdmanis, J. F. Whitaker, G. A. Mourou, F. W. Smith, A. R. Calawa: Appl. Phys. Lett. **59**, 3276–3278 (1991)
- 45.43 E. S. Harmon, M. R. Melloch, J. W. Woodall, D. D. Nolte, N. Olsuka, C. L. Chang: Appl. Phys. Lett. **63**, 2248–2250 (1993)
- 45.44 S. D. Benjamin, A. Othonos, P. W. E. Smith: Electron. Lett. **30**, 1704–1706 (1994)
- 45.45 P. W. E. Smith, S. D. Benjamin, H. S. Loka: Appl. Phys. Lett. **71**, 1156–1158 (1997)
- 45.46 M. Kawase, E. Garmire, H. C. Lee, P. D. Dapkus: IEEE J. Quantum Electron. **30**, 981–988 (1994)
- 45.47 L. Brzozowski, E. H. Sargent, A. SpringThorpe, M. Extavour: Appl. Phys. Lett. **82**, 4429–4431 (2003)
- 45.48 D. A. B. Miller, D. S. Chemla, D. J. Eilenberger, P. W. Smith, A. C. Gossard, W. T. Tsang: Appl. Phys. Lett. **41**, 679–681 (1982)
- 45.49 S. Schmitt-Rink, D. S. Chemla, D. A. B. Miller: Phys. Rev. B **32**, 6601–6609 (1985)
- 45.50 Y. H. Lee, A. Chavez-Pirson, S. W. Koch, H. M. Gibbs, S. H. Park, J. Morchange, A. Jeffery, N. Peyghambarian, J. Banyai, A. C. Gossard, W. Wiegmann: Phys. Rev. Lett. **57**, 2446–2449 (1986)
- 45.51 F. Stern: Phys. Rev. **133**, A1653–A1664 (1964)
- 45.52 A. M. Fox, A. C. Maciel, J. F. Ryan, M. D. Scott: Nonlinear excitonic optical absorption in GaInAs/InP quantum wells
- 45.53 A. M. Fox, A. C. Maciel, M. G. Shorthose, J. F. Ryan, M. D. Scott, J. I. Davies, J. R. Riffat: Nonlinear excitonic optical absorption in GaInAs/InP quantum wells
- 45.54 L. Brzozowski, E. H. Sargent, A. SpringThorpe, M. Extavour: Appl. Phys. Lett., 4429–4431 (2003)
- 45.55 L. Brzozowski, E. H. Sargent, A. SpringThorpe, M. Extavour: Virtual J. Ultrafast Sci. **2**(7) (2003)
- 45.56 P. W. Joudawlkis, D. T. McInturff, S. E. Ralph: Appl. Phys. Lett. **69**, 4062–4064 (1996)
- 45.57 R. V. Penty, H. K. Tsang, I. H. White, R. S. Grant, W. Sibert, J. E. A. Whiteaway: Electron. Lett. **27**, 1447–1449 (1991)
- 45.58 I. E. Day, P. A. Snow, R. V. Penty, I. H. White, R. S. Grant, G. T. Kennedy, W. Sibbett, D. A. O. Davies, M. A. Fisher, M. J. Adams: Appl. Phys. Lett. **65**, 2657–2659 (1994)
- 45.59 M. A. Fisher, H. Wickes, G. T. Kennedy, R. S. Grant, W. Sibbett: Electron. Lett. **29**, 1185–1186 (1993)
- 45.60 D. A. O. Davies, M. A. Fisher, D. J. Elton, S. D. Perin, M. J. Adams, G. T. Kennedy, R. S. Grant,

- P. D. Roberts, W. Sibbett: *Electron. Lett.* **29**, 1710–1711 (1993)
- 45.61 C. Aversa, J. E. Sipe, M. Sheik-Bahae, E. W. V. Stryland: *Phys. Rev. B* **24**, 18073–18082 (1994)
- 45.62 M. J. Shaw, M. Jaros: *Phys. Rev. B* **47**, 1620–1623 (1993)
- 45.63 K. S. Bindra, A. K. Kar: *Appl. Phys. Lett.* **79**, 3761–3763 (2001)
- 45.64 M. Sheik-Bahae, D. J. Hagan, E. W. Van Stryland: *Phys. Rev. Lett.* **65**, 96–99 (1990)
- 45.65 N. J. Long: *Angew. Chem., Int. Ed.* **34**, 21–38 (1995)
- 45.66 I. Liakatas, C. Cai, M. Bösch, C. B. M. Jäger, P. Günter: *Appl. Phys. Lett.* **76**, 1368–1370 (2000)
- 45.67 R. Rangel-Rojo, S. Yamada, H. Matsuda, D. Yankelevich: *Appl. Phys. Lett.* **72**, 1021–1023 (1998)
- 45.68 R. Rangel-Rojo, S. Yamada, H. Matsuda, H. Kasai, H. Nakanishi, A. K. Kar, B. S. Wherrett: *J. Opt. Soc. Am. B* **203**, 2937–2945 (1998)
- 45.69 R. Rangel-Rojo, S. Yamada, H. Matsuda, H. Kasai, Y. Komai, S. Okada, H. Oikawa, H. Nakanishi: *Jap. J. Appl. Phys.* **38**, 69–73 (1999)
- 45.70 R. Rangel-Rojo, H. Matsuda, H. Kasai, H. Nakanishi: *J. Opt. Soc. Am.* **17**, 1376–1382 (2000)
- 45.71 E. Van Keuren, T. Wakebe, R. Andreaus, H. Möhwald, W. Schrof, V. Belov, H. Matsuda, R. Rangel-Rojo: *J. Opt. Soc. Am. B* **203**, 2937–2945 (1998)
- 45.72 R. Rangel-Rojo, L. Stranges, A. K. Kar, M. A. Mendez-Rojas, W. H. Watson: *Opt. Commun.* **203**, 385–391 (2002)
- 45.73 L. Demenicis, A. S. L. Gomes, D. V. Petrov, C. B. de Araújo, C. P. de Molo, C. G. dos Santos, R. Souto-Major: *J. Opt. Soc. Am. B* **14**, 609–614 (1997)
- 45.74 C. Sauteret, J. P. Hermann, R. Frey, F. Pradère, J. Duclung, R. H. Baughman, R. R. Chance: *Phys. Rev. Lett.* **36**, 956–959 (1976)
- 45.75 G. P. Agrawal, C. Cojan, C. Flytzanis: *Phys. Rev. B* **17**, 776–789 (1978)
- 45.76 S. R. Marder, B. Kippelen, A. Y. Jan, N. Peyghambarian: *Nature* **388**, 845–951 (1997)
- 45.77 B. L. Lawrence, M. Cha, J. U. Kang, W. Torruellas, G. Stegeman, G. Baker, J. Meth, S. Etemad: *Electron. Lett.* **30**, 447–448 (1994)
- 45.78 B. L. Lawrence, W. Torruellas, M. C. G. Stegeman, J. Meth, S. Etemad, G. Baker: *Electron. Lett.* **30**, 447–448 (1994)
- 45.79 F. Yoshino, S. Polyakov, L. Friedrich, M. Liu, H. Shim, G. I. Stegeman: *J. Nonlinear Opt. Phys. Mater.* **9**, 95–104 (2000)
- 45.80 F. Yoshino, S. Polyakov, M. Liu, G. Stegeman: *Phys. Rev. Lett.* **91**, 063901–1–063901–4 (2003)
- 45.81 S. Wang, W. Huang, R. Liang, Q. Gong, H. Li, H. Chen, D. Qiang: *Phys. Rev. B* **63**, 153408(1–4) (2001)
- 45.82 B. L. Yu, H. P. Xia, C. S. Zhu, F. X. Gan: *Appl. Phys. Lett.* **81**, 2701–2703 (2002)
- 45.83 Q. Chen, L. Kuang, E. H. Sargent, Z. Y. Wang: *Appl. Phys. Lett.* **83**, 2115–2117 (2003)
- 45.84 F. W. Wise: *Accounts Chem. Res.* **33**, 773–780 (2000)
- 45.85 L. Banyai, S. W. Koch: *Phys. Rev. Lett.* **57**, 2722–2724 (1986)
- 45.86 S. Schmitt-Rink, D. A. B. Miller, D. S. Chemla: *Phys. Rev. B* **35**, 8113–8125 (1987)
- 45.87 M. A. Hines, G. D. Scholes: *Synthesis of colloidal PbS nanocrystals with size-tunable NIR emissions*, submitted
- 45.88 G. Wang, K. Guo: *Physica B* **315**, 234–239 (2001)
- 45.89 P. T. Guerreiro, S. Ten, N. F. Borrelli, J. Butty, G. E. Jabbour, N. Peyghambarian: *Appl. Phys. Lett.* **71**, 1595–1597 (1997)
- 45.90 K. Wundke, S. Pötting, J. Auxier, A. Schülzgen, N. Peyghambarian, N. F. Borrelli: *Appl. Phys. Lett.* **76**, 10–12 (2000)
- 45.91 A. M. Malyarevich, V. G. Savitski, P. V. Prokoshin, N. N. Posonov, K. V. Yumashev, E. Raaben, A. A. Zhilin: *J. Chem. Phys.* **78**, 1543–1551 (1983)
- 45.92 D. Cotter, M. C. Burt, R. J. Manning: *Phys. Rev. Lett.* **68**, 1200–1203 (1992)
- 45.93 H. P. Li, B. Liu, C. H. Kam, Y. L. Lam, W. X. Que, L. M. Gan, C. H. Chew, G. Q. Xu: *Opt. Mater.* **14**, 321–327 (2000)
- 45.94 B. Liu, H. Li, C. H. Chew, W. Que, Y. L. Lam, C. H. Kam, L. M. Gan, G. Q. Xu: *Mater. Lett.* **51**, 461–469 (2001)
- 45.95 B. Liu, C. H. Chew, L. M. Gan, G. Q. Xu, H. Li, Y. L. Lam, C. H. Kam, W. X. Que: *Mater. Lett.* **51**, 461–469 (2001)
- 45.96 L. Yang, K. Becker, F. M. Smith, R. H. Margruder III, R. F. Haglund Jr., L. Yang, R. Dorsinville, R. R. Alfano, R. A. Zühr: *Size dependence of the third-order susceptibility of copper nanocrystals investigated by four-wave mixing*
- 45.97 K. Uchida, S. Kaneko, S. Omi, C. Hata, H. Tanji, Y. Asahara, A. J. Ikushima, T. Tokizaki, A. Nakamura: *J. Opt. Soc. Am.* **11**, 1236–1243 (1994)
- 45.98 H. Inouye, K. Tanaka, I. Tanahashi, T. Hattori, H. Nakatsuka: *Jap. J. Appl. Phys.* **39**, 5132–5133 (2000)
- 45.99 D. Ricard, P. Roussignol, C. Flytanis: *Opt. Lett.* **10**, 511–513 (1985)
- 45.100 D. V. Petrov: *Opt. Commun.* **13**, 102–106 (1996)
- 45.101 C. Bosshard: *Adv. Mater.* **5**, 385–397 (1996)
- 45.102 G. I. Stegeman, D. J. Hagan, L. Torner: *Opt. Quantum Electron.* **28**, 1691–1740 (1996)

Novel Materials

Part E

Part E Novel Materials and Selected Applications

46 Solar Cells and Photovoltaics

Stuart Irvine, Gwynedd, UK

47 Silicon on Mechanically Flexible Substrates for Large-Area Electronics

Peyman Servati, Waterloo, Canada
Arokia Nathan, Waterloo, Canada

48 Photoconductors for X-Ray Image Detectors

M. Zahangir Kabir, Montreal, Canada
Safa Kasap, Saskatoon, Canada
John Rowlands, Toronto, Canada

49 Phase-Change Optical Recording

Alexander Kolobov, Ibaraki, Japan
Junji Tominaga, Tsukuba, Japan

50 Carbon Nanotubes and Bucky Materials

Mark Baxendale, London, UK

51 Magnetic Information-Storage Materials

Larry Comstock, San Jose, USA

52 High-Temperature Superconductors

Rainer Wesche, Lausanne, Switzerland

53 Molecular Electronics

Michael Petty, Durham, UK

54 Organic Materials for Chemical Sensing

Asim Kumar Ray, London, UK

55 Packaging Materials

Darrel Frear, Tempe, USA

Solar Cells and Photovoltaics

Photovoltaic solar cells are gaining wide acceptance for producing clean, renewable electricity. This has been based on more than 40 years of research that has benefited from the revolution in silicon electronics and compound semiconductors in optoelectronics. This chapter gives an introduction into the basic science of photovoltaic solar cells and the challenge of extracting the maximum amount of electrical energy from the available solar energy. In addition to the constraints of the basic physics of these devices, there are considerable challenges in materials synthesis. The latter has become more prominent with the need to reduce the cost of solar module manufacture as it enters mainstream energy production. The chapter is divided into sections dealing with the fundamentals of solar cells and then considering five very different materials systems, from crystalline silicon through to polycrystalline thin films. These materials have been chosen because they are all in production, although some are only in the early stages of production. Many more materials are being considered in research and some of the

46.1	Figures of Merit for Solar Cells	1096
46.2	Crystalline Silicon	1098
46.3	Amorphous Silicon	1100
46.4	GaAs Solar Cells	1101
46.5	CdTe Thin-Film Solar Cells	1102
46.6	CuInGaSe₂ (CIGS) Thin-Film Solar Cells	1103
46.7	Conclusions	1104
	References	1105

more exciting, polymer and dye-sensitised cells are mentioned in the conclusions. However, there is insufficient space to give these very active areas of research the justice they deserve. I hope the reader will feel sufficiently inspired by this topic to read further and explore one of the most exciting areas of semiconductor science. The need for high-volume production at low cost has taken the researcher along paths not normally considered in semiconductor devices and it is this that provides an exciting challenge.

Photovoltaic (PV) devices are a method of converting radiant solar energy into electrical energy. Most of our energy sources, including fossil fuels, hydroelectric and wind power actually come from solar radiation but are indirect conversions into electricity. Another class of solar energy conversion is the heating of water in solar thermal panels. Although the conversion efficiency can be high, they do not generate the thermal energy necessarily where and when it is needed, so storage is required. Direct generation of electric energy is attractive because it is a versatile energy form, rapidly converted into heat, mechanical or light energy. Photovoltaic energy is the main source of energy in the rapidly expanding satellite market with high-efficiency photovoltaic modules producing more than 1 kW of power. Terrestrial applications are also rapidly growing with an estimated total installed capacity worldwide over 1 GW in 2004, increasing annually by 30–40%. This, how-

ever, is tiny compared with the total amount of electrical energy consumed each year, around 0.1% of primary energy demand. However, solar energy is very attractive as it is completely non-polluting and can help to reduce the amount of fossil fuels that we burn to generate electricity. World CO₂ emissions have grown by 8% since 1990. Any contribution from non-fossil-fuel alternatives such as solar energy will help to reduce this annual burden of CO₂ emissions that is now a widely accepted cause of global warming. So how much solar energy is available for conversion to electricity? The total solar energy falling on the Earth's surface each year is huge and 10 000 times the current consumption of energy. We only need to capture a tiny fraction of this to make a major contribution to our electricity supply but this will mean incorporating solar electric panels into most buildings; to achieve this it will need to be much cheaper to compete with current fossil-fuel electricity generation.

The range of applications outlined here place different demands on the design of the photovoltaic module and hence the materials solutions may well be different. Cost has already been mentioned as paramount for terrestrial power requirements, but for space applications resistance to ionising radiation is important. This section will consider the materials implications of competing technologies and how well they meet the criteria for different applications. The first section will deal with how a photovoltaic cell operates and introduce the

figures of merit that are used to compare photovoltaic cells and relate performance to theoretical performance. Solar-cell technology is promoted as a clean fuel technology which does not introduce CO_2 and pollutants into the atmosphere. However, for this to be a truly environmentally friendly technology, the manufacturing and eventual disposal will also have to be environmentally safe and this depends on the materials used and the fabrication technologies. These factors will be considered for each material system.

46.1 Figures of Merit for Solar Cells

In this section we will consider the operation of a photovoltaic cell and the significant parameters that characterise its performance. The objective of a photovoltaic cell is to capture as much of the solar energy as possible and convert this into electrical energy. The solar energy reaching the Earth's atmosphere fits a black-body distribution for a body at 5800 K. This spectrum becomes highly structured, particularly in the infrared part of the spectrum, by absorption bands due to atmospheric gases. By the time the solar radiation reaches the Earth's surface, it no longer fits a black-body distribution [46.1]. The shift in the spectral distribution will obviously affect the efficiency of absorption of the solar radiation, particularly considering that all semiconductor materials will display a cut-off wavelength dictated by the band gap of the semiconductor. It is necessary to specify the atmospheric absorption when quoting efficiency, as the depth of atmosphere that the solar radiation has passed through will affect both the spectral distribution and the total amount of energy. The measurement used is air mass (AM) and is defined as zero for solar radiation outside the atmosphere and 1 for radiation reaching the ground when the sun is at its zenith. For shallower angles the solar radiation has to penetrate a larger depth of atmosphere and the AM is therefore going to be greater than 1. For space applications AM 0 is the appropriate condition and for terrestrial applications (depending on the position on the Earth) AM 1.5 is a typical value quoted, which corresponds to a solar angle of 45° . All this assumes that there is no cloud cover, which will further reduce the integrated intensity and modify the solar spectrum. Although cloud cover will reduce the amount of solar energy available for conversion, useful amounts of electricity can still be generated.

The available solar energy also decreases from 1353 W/m^2 outside the atmosphere (AM 0) to 925 W/m^2 for AM 1, when the sun is directly over-

head. In practice, for terrestrial applications the available solar energy is considerably less than this where, in general, $\text{AM} > 1$ and is further reduced by cloud cover.

A photovoltaic cell is basically a diode with a photogenerated current. A band diagram is shown schematically in Fig. 46.1. Absorption of radiation can occur on both sides of the junction, creating minority carriers that can diffuse towards the junction. A photocurrent is generated if the minority carriers can drift across the junction without recombination. In practice, the junction is shallow and absorption will occur predominantly on one side where there is a greater depth of absorbing material. There are also heterojunction p-n devices where less absorption will occur in the wider-band-gap layer, as shown in Fig. 46.2. So, one side of the junction is considered to be the absorber layer and it is the spectral absorption characteristics of this layer that

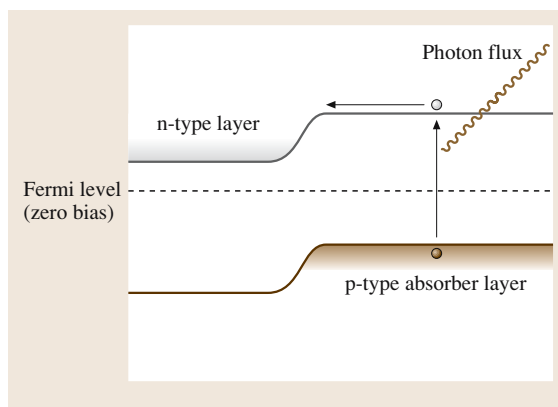


Fig. 46.1 Schematic of an energy-band diagram for a p-n junction solar cell showing photoabsorption to create an electron-hole pair and diffusion of the electron towards the junction

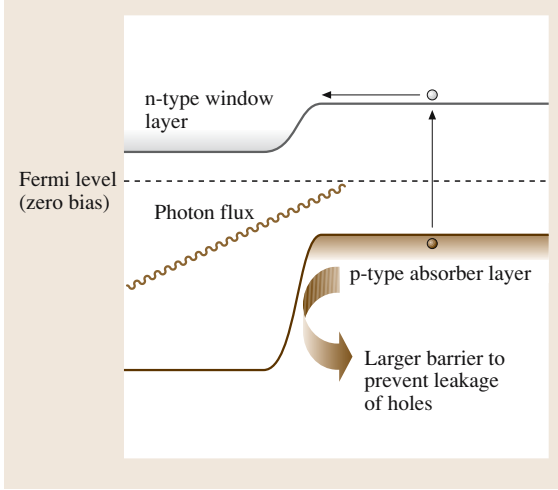


Fig. 46.2 Band diagram for a heterojunction PV cell showing illumination through the wide-band-gap window layer

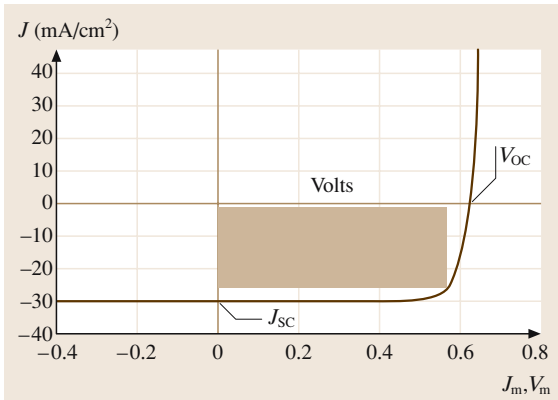


Fig. 46.3 Ideal J - V characteristic for a photovoltaic cell, according to (46.1) with J_s equal to 30 mA/cm^2 . The parameters J_{sc} and V_{oc} are indicated on the graph. The maximum extracted power is shown as the shaded area

will determine the maximum absorption possible from the available solar radiation.

An ideal diode characteristic for an illuminated cell is shown in Fig. 46.3. In the dark the J - V plot would go through the origin and, as the light intensity increases, the short-circuit current becomes increasingly negative, indicating the presence of a photogenerated current. The equation for the J - V characteristic of this ideal device is:

$$J = J_s \left(e^{qV/k_B T} - 1 \right) - J_L, \quad (46.1)$$

where J_s is the saturation current in reverse bias under zero illumination, q is the charge on the carrier, V

is the applied voltage, k_B is Boltzmann's constant, T is the temperature of the cell and J_L is the photogenerated current. In the ideal cell this is equal to the short-circuit current, indicated as J_{sc} on the J - V curve in Fig. 46.3. Power can be extracted from the device in the $+V$, $-J$ quadrant of the J - V plot and the load will determine the operating point in this quadrant. The power is determined from the product JV at this operating point, shown graphically in Fig. 46.3. The maximum power will correspond to the operating point that will give the largest JV area on this graph. For the ideal diode characteristics given in (46.1), this maximum power is given by the following (46.2):

$$\begin{aligned} P_m &= I_m V_m \\ &= I_L \left[V_{oc} - \frac{k_B T}{q} \ln \left(1 + \frac{q V_m}{k_B T} \right) - \frac{k_B T}{q} \right] \\ &= I_L (E_m/q), \end{aligned} \quad (46.2)$$

where E_m is the maximum energy that can be extracted per photon and depends on the band parameters for the semiconductor absorber layer, which determine V_{oc} and V_m . These parameters are marked on the J - V plot in Fig. 46.3. We now have two fundamental parameters which will limit the efficiency of the cell:

- The fraction of solar photons absorbed in the cell.
- The electrical energy created per photon.

The first factor can be calculated by integrating over the solar spectrum for the appropriate AM number and including the cut-off wavelength of the semiconductor absorber layer.

$$\eta_{\text{abs}} = \frac{\int_{E_g}^{\infty} n_E(E) dE}{\int_0^{\infty} n_E(E) dE}. \quad (46.3)$$

This is shown graphically in Fig. 46.4 for AM 1.5. Any photons with energy less than the band gap will not be absorbed and will not contribute to the photocurrent.

The second efficiency factor mentioned above implies that not all the photon energy will be converted into electrical energy, even if one photon absorbed constitutes one minority carrier crossing the junction. The electrical energy per carrier is given by the factor E_m in (46.2), so the maximum power of the device is the product of the absorption rate of photons and the mean electrical energy created per photon. This product is represented by the inner shaded area of the solar spectrum shown in Fig. 46.4. The difference between curve 1 and

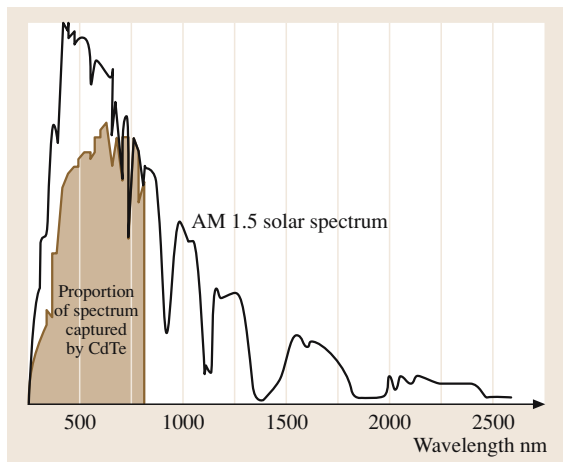


Fig. 46.4 Graphical representation of the maximum energy that can be extracted from a CdTe solar cell with a band-gap energy of 1.45 eV

curve 2 is simply the energy lost per photon because not all the photon energy is converted into electrical energy. Different semiconductor materials will have different efficiencies primarily because of different values for the band gap. The ideal value for E_m will track the band gap, so for narrower-band-gap materials there will be a larger proportion of photons absorbed but less electrical energy per photon. The function of efficiency for semiconductors with different band gap, taken from curve 2 in Fig. 46.4 is plotted in Fig. 46.5 and shows that the optimum efficiency occurs for semiconductors with a band gap in the near-infrared region, around 1.5 eV. This represents the best compromise between absorption of solar radiation and transferring the optimum amount of energy per photon into electrical energy.

The maximum efficiency is predicted for Si, InP, GaAs and CdTe, which are in the region of 30% for AM 1.5 irradiation (1 sun illumination). This means that the very best conversion efficiency for a single-junction cell is approximately one third of the available solar energy. In practice, photovoltaic cells have efficiencies considerably less than this due to optical reflections, poor junction characteristics and carrier recombination. These are materials issues that are often traded off against production costs, e.g. using polycrystalline

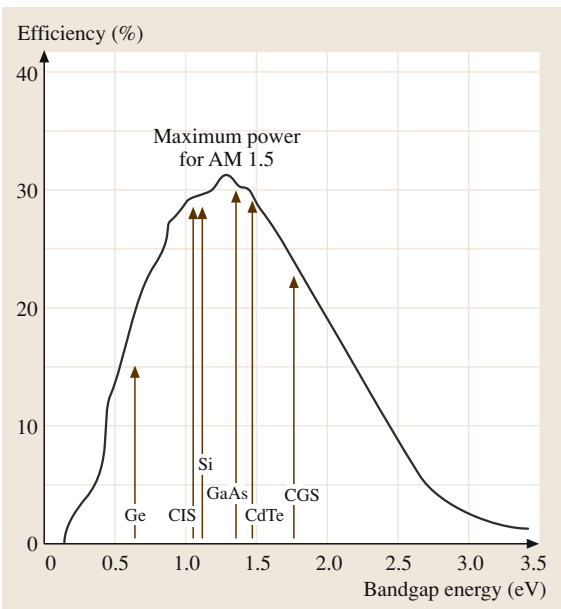


Fig. 46.5 Plot of ideal efficiency against band-gap energy for a single-junction cell for AM 1.5 illumination conditions, (after Henry [46.1])

rather than single-crystal material. Conversely, higher conversion efficiencies can be achieved using multiple junctions which are more expensive but attractive for space applications and when used in combination with solar concentrators, so less surface area of the expensive multijunction cell is required. Other factors that can influence the choice of photovoltaic materials include the following:

- absorption coefficient,
- contact resistance,
- abundance of raw materials,
- toxicity of materials,
- stability of materials and junctions,
- radiation resistance.

These factors will be considered in the following examples of different photovoltaic systems to assess the merits of different materials. It is probably true to say that there is not one ideal material but different applications can make one material more attractive than another.

46.2 Crystalline Silicon

Over 80% of the current world production of solar modules are made from either single-crystal or

multigrain silicon. This is the most mature of the photovoltaic materials and has benefited considerably

from the size of the silicon semiconductor industry. This has ensured a ready supply of material and processing tools suitable for large-scale production. However, crystalline silicon does suffer from a fundamental disadvantage in that the band gap of silicon is indirect, which means that the absorption coefficient is much lower than a direct-band-gap semiconductor such as CdTe. In practice this means that a much thicker piece of material is needed to absorb all the solar radiation greater than the band-gap energy. This requires wafers of silicon thicker than $100\text{ }\mu\text{m}$ and means that this material is not suitable for thin-film technology. The highest-performance solar electric modules are made from wafers of single-crystal silicon cut from Czochralski-grown crystals, up to 30 cm in diameter. This was originally developed on the back of the silicon electronics industry, but the rapid increase in production volume of PVs is now driving production of silicon crystals. The conversion efficiency for single-crystal PV modules is around 17% with the potential for increases in the near future to 20%.

Multicrystalline silicon involves the relatively cheap path of casting silicon ingots that are not seeded but produce a random grain size of the order of 1 cm across. The ingots are cast in blocks, larger than 100 kg, and then sliced into wafers $300\text{-}\mu\text{m}$ thick and with an area of up to 20 cm^2 . The grain boundaries cause these wafers to be mechanically weaker than the single-crystal wafers and they will typically be thicker. This causes some trade off in price. Each cell can be expected to contain grain boundaries, so loss of photogenerated charge at

the grain boundary can cause some loss of efficiency. Typical module efficiencies for multicrystalline silicon are currently around 15%.

The junction is formed by phosphorus implantation to form a p-n homojunction and contacted by printing of a metal grid, usually of a Ni–Au alloy. Patterning of the surface prior to implantation and contacting has achieved the highest efficiency cells by improving light collection [46.2]. Another recent innovation that has contributed to higher efficiencies is the ‘V’-groove or buried junction. This is shown schematically in Fig. 46.6 and entails the grooving of the p-type silicon substrate, implantation of phosphorus to create a buried junction and filling with metal contact alloy. This helps to improve the collection efficiency of the cell, particularly in the blue part of the spectrum. These manufacturing steps have to handle large volumes and be cheap. Cheaper alternatives to ion implantation involve thermal diffusion from a screen-printed paste or spin-on glass [46.3].

The main disadvantage of silicon is that the absorption coefficient is low because it is an indirect-band-gap semiconductor ($2 \times 10^3\text{ cm}^{-1}$ for Si compared with $1 \times 10^5\text{ cm}^{-1}$ for CdTe in the green part of the spectrum). This means that the amount of material needed to absorb the solar radiation is greater than for a direct-band-gap semiconductor. The absorption can be improved by having a reflecting back contact that, if it is roughened, will increase the mean path length of the back reflected light in the silicon cell. This is particularly important for thin-film silicon devices where the amount of material is more of an issue, although the thickness of $30\text{--}100\text{ }\mu\text{m}$ can be compared with the much thinner layers used for direct-band-gap semiconductors (see CdTe and GaAs).

Thin-film polycrystalline silicon is grown by ribbon casting techniques. The pulling speeds are in the range $10\text{--}1800\text{ cm/min}$, and this represents a cheap method for the production of solar cell material but the conversion efficiencies are currently low. One way of increasing the absorption coefficient is to deposit a film of amorphous silicon (a-Si), which has a similar band gap to crystalline silicon but is a direct-band-gap material. Amorphous silicon has led the way in cheap thin-film solar cells but suffers the disadvantages of low efficiency ($< 10\%$) and poor long-term stability. These cells will be discussed in greater detail in the next section.

Crystalline silicon solar cells, over the past 10 years, have approximately doubled in efficiency and this has been achieved by a combination of the

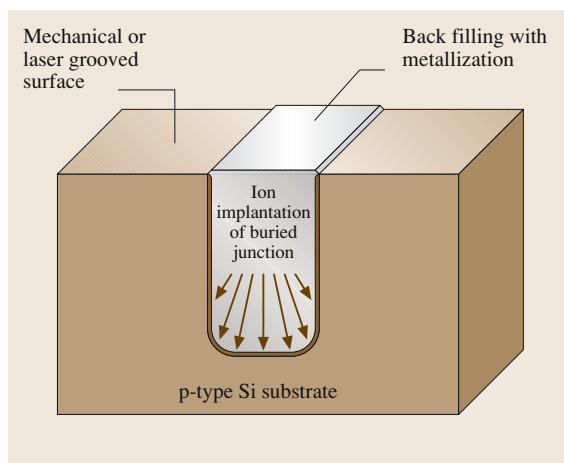


Fig. 46.6 Schematic of buried-contact technology for Si solar cells

following:

- Improved material quality, leading to improved minority-carrier diffusion length.
- Improved V_{oc} and fill factor through emitter and base doping and contact optimisation.
- Improved J_{sc} through diffusion length improvement using phosphorus gettering, hydrogen passivation and buried contacts.
- Surface passivation and contact grid optimisation.

All these aspects involve materials issues, either directly associated with the quality of the silicon or with contacting and passivation. It is also important to avoid degradation of the cell and the final encapsulation must avoid exposure of the cell to water. Multicrystalline sil-

icon can be passivated with silicon nitride deposited by plasma-assisted CVD from SiH_4 and NH_3 to reduce surface recombination [46.4].

Silicon solar modules are becoming cheaper to produce, nontoxic and stability in a non-radiation environment is good. Multicrystalline silicon is not so suitable for space applications because of the combination of thick absorber layers, requiring greater weight per unit area and the sensitivity of the cell to degrade in a high-radiation environment. However, single-crystal silicon competes well with GaAs cells in this market. Single-crystal modules for terrestrial applications produce the highest output powers currently available with a Sharp 1.3 m² module giving a peak output power of 185 W.

46.3 Amorphous Silicon

Amorphous silicon offers the potential for a cheap production technology for terrestrial photovoltaic solar cells. The amorphous state displays different physical properties to the crystalline with a modified band structure. One consequence of this is that the absorption coefficient in the green part of the visible spectrum is a factor of 10 higher at $2 \times 10^4 \text{ cm}^{-1}$, which makes it more suitable for thin-film applications. The amorphous structure leaves dangling bonds which pin the Fermi level and would normally prevent doping of the material. This is overcome by the inclusion of hydrogen, which passivates these dangling bonds, so the material is referred to as a-Si:H. The Si can also be alloyed with Ge, C, and N in a glow-discharge evaporator. These alloys are particularly useful for multijunction devices used for increasing the quantum efficiency. In the laboratory single-junction cells have efficiency around 8% with multijunction cells going up to 20%. In production an a-Si module has efficiency around 7%, considerably less than the crystalline materials.

The most common method for producing a-Si:H for photovoltaics is by plasma-enhanced CVD from SiH_4 mixtures. The films are deposited onto a textured conducting oxide such as indium–tin oxide (ITO) which provides the electrical contact and increases the average light path in the absorber layer to increase absorption. The device structure is a p-i-n with absorption taking place in the middle (insulator) layer, which is only 0.5 μm thick. The p and n layers can be deposited by adding B_2H_6 or PH_3 respectively to the plasma.

One of the major disadvantages of a-Si:H is the instability and long-term deterioration under light il-

lumination. This is caused by the rearrangement of dangling bonds, often associated with rearrangement of hydrogen atoms close to weak or dangling bonds. The energy comes from nonradiative bimolecular reactions and hence depends on the illumination intensity (for further details see the review by Bloss et al. [46.5]). In practice, this causes a downward drift in efficiency with time which can be as much as 2% in 100 h. The current cost is cheaper than crystalline silicon (3 US\$/Wp) but with the potential to considerably reduce cost to 0.7 US\$/Wp [46.6]. One of the main technical challenges is to maintain stabilised efficiency above 10%. The relatively high cost of PV solar modules compared with more conventional sources of energy can only become attractive if it can operate efficiently over a long period of time. This means that the modules must remain efficient over a period of 20 years. Shorter operating life times would have to be mitigated by much lower production cost.

In conclusion, a-Si:H is a low-cost technology for terrestrial applications and is finding its way into low-power applications such as small-scale stand-alone systems. With improved efficiency and stability s-Si has the potential to capture a significant proportion of the terrestrial market but the lower production costs are not sufficient to offset this at present. A typical a-Si module, with an area of 0.8 m², will have a peak output power of 40 W. The temperatures used in processing these modules is lower than the high temperatures needed to melt silicon for the crystalline silicon modules which means that the energy payback times could be just a few months.

46.4 GaAs Solar Cells

The III–V semiconductors have advantages over silicon due to their direct-band-gap photon absorption, with an absorption coefficient in the green of $8 \times 10^4 \text{ cm}^{-1}$. This means that, for example, GaAs can theoretically yield over 30% efficiency (AM 1.5) for an absorber layer thickness of just 1–2 μm , compared with a hundred times this for crystalline silicon. The efficiency, stability and thin-film deposition technology make these cells attractive for space applications. From Fig. 46.5 it is clear that the band gap of GaAs is well matched to the optimum for maximum efficiency. The III–V semiconductors also offer the flexibility of alloying to change the band gap and tune the response of the photovoltaic junction. In addition, heterojunctions can be formed and multijunction solar cells can be produced to convert more of the solar spectrum into electricity and thus exceed the theoretical limits set by the single-junction cells. For example, in the laboratory a GaAs/GaSb tandem solar cell has been reported with 35.6% efficiency [46.8]. Higher efficiencies are possible with three and even four junctions to capture more of the infrared that would otherwise not be absorbed. The Ge substrate, which is closely lattice matched with GaAs, can be used as a narrow-band-gap absorber to capture the radiation that passes through the GaInP₂/GaAs cell [46.7]. The top cell has an absorber of GaInP₂, which has a band gap of 1.9–2 eV, and will capture the visible part of the spectrum without too much voltage loss. The next cell is GaAs, which has a band gap of 1.42 eV, and will capture the red to infrared part of the spectrum. The bottom junction is formed in the Ge substrate, which has a band gap of 0.67 eV and will capture the light further into the infrared. The triple-junction cell, in production, can yield 27% (AM 0) conversion efficiency [46.7].

The layers of these multijunction cells are grown epitaxially onto a single-crystal Ge substrates. As with crystalline silicon, high crystalline quality is needed to obtain high-efficiency cells; polycrystalline GaAs does not work due to grain-boundary conduction that reduces the available photocurrent. This restricts the photovoltaic applications to high-quality epitaxial material and GaAs is not suitable for cheap thin-film photovoltaics on glass substrates. The topic of thin-film polycrystalline photovoltaics will be considered in the next section on CdTe solar cells. Epitaxial growth avoids minority-carrier recombination at grain boundaries, by avoiding their formation, but the cost of the substrates is inherently higher than the glass or ceramic substrates used for thin-film devices. However, defects must also

be avoided at the junction and interfaces between layers of different composition and this imposes the constraint that the heterostructures should be lattice matched. This restricts the choice of alloys to those that are either lattice matched or that the thickness and mismatch are such that the film is strained. The simplest structures use AlGaAs, which has a close lattice match to GaAs over the entire composition range, as a window and passivation layer. GaAs is also closely matched to Ge, offering a choice of substrates. The GaAs or Ge substrate is the narrowest-band-gap part of the structure, so the device has to be front-side illuminated with a grid of metal contacts on the top surface to contact the cells. The design of the cell structure needs to have the wider-band-gap layers further from the substrate (last to be grown) to allow the longer wavelengths to penetrate to the narrow-band-gap absorber layer. This places further constraints on the design of these epitaxial structures.

The Ge and GaAs substrates have a similar lattice parameter to AlGaAs, which is used as a passivation layer for both n-type and p-type GaAs layers. Other III–

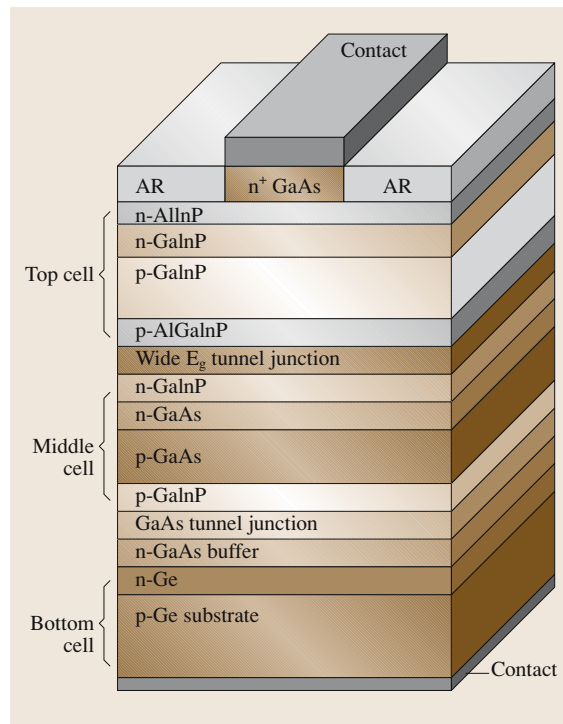


Fig. 46.7 Schematic of a triple-junction GaAs solar cell [46.7], showing the positions of the three junctions

V compounds that can be used include GaSb, which is sensitive to the near infrared out to $1.8\text{ }\mu\text{m}$ and will therefore absorb the radiation that is transmitted through the GaAs cell, as an alternative to the Ge-bottom cell. The alloy GaInP also lattice matches for the 50% Ga/In mix and is useful as a wider-band-gap junction. GaSb is also attractive for thermo-photovoltaics where the solar radiation is converted into heat, which is then absorbed by the narrow-band-gap GaSb device. More recently, the narrow-band-gap quaternary GaInNAs has been investigated [46.9] as a narrow-band-gap absorber layer and can be inserted between the GaAs layer and the Ge substrate to give a small boost in voltage.

The III–V photovoltaic structures are grown by metalorganic vapour-phase epitaxy (MOVPE), which gives excellent control over alloy composition and doping concentration. An example of a multijunction cell is shown

in Fig. 46.7. The structures tend to be complex because a tunnel junction is needed for current to flow between the two (series) photovoltaic cells, otherwise the connection between the two would be a high-impedance reverse-bias junction. It is also important for the device design to match the currents between the two (or more) junctions so that they can both operate at near-optimum conditions.

GaAs solar cells are rapidly increasing their share of the space market to power satellites. Their efficiency is high, stability good but the single-crystal substrates increase the cost. There may be toxicity issue for the disposal of solar modules if GaAs was used widely for terrestrial applications. The most attractive terrestrial application would be in concentrators where the solar radiation is optically concentrated onto the PV cells, so the collection area can be much greater than the area of expensive PV modules.

46.5 CdTe Thin-Film Solar Cells

The band gap of CdTe of 1.45 eV at room temperature makes it another semiconductor that is close to the theoretical optimum for efficient conversion of solar radiation (see Fig. 46.5) into electrical power. The absorption coefficient is $> 5 \times 10^4\text{ cm}^{-1}$ for photon energy greater than the band gap, allowing efficient collection for only $2\text{-}\mu\text{m}$ -thick films, similar to GaAs. CdTe, like GaAs is also a direct-band-gap semiconductor but is from the II–VI family of semiconductor compounds. Although the theoretical efficiency for the CdS/CdTe photovoltaic cell is $\approx 30\%$ at AM 1.5, the highest values reported are nearly half this maximum and are less than 10% at AM 1.5 for modules in production. The attraction of CdTe compared with GaAs is that these cells can be made from polycrystalline thin films on glass substrates, thus avoiding the need for expensive single-crystal substrates. A further advantage in using glass substrates is that illumination of the photovoltaic cell occurs through the substrate rather than from the top face, so the substrate becomes the window for the cell. The front contact is made with a transparent conducting oxide (TCO) such as ITO, as shown in Fig. 46.8, and this avoids the need for an opaque grid of metal contacts. This approach of using a TCO on glass substrate as the window is called a superstrate. The TCO has to have high optical transmission as well as a metal-like electrical conductivity. A typical spreading resistance would be $10\text{ }\Omega$ per square.

The CdTe solar cell is made from a heterojunction between the wider-band-gap CdS and the CdTe absorber layer. CdS is often referred to as the window layer; it has a band gap of 2.4 eV allowing most of the visible spectrum to pass with very little absorption. The

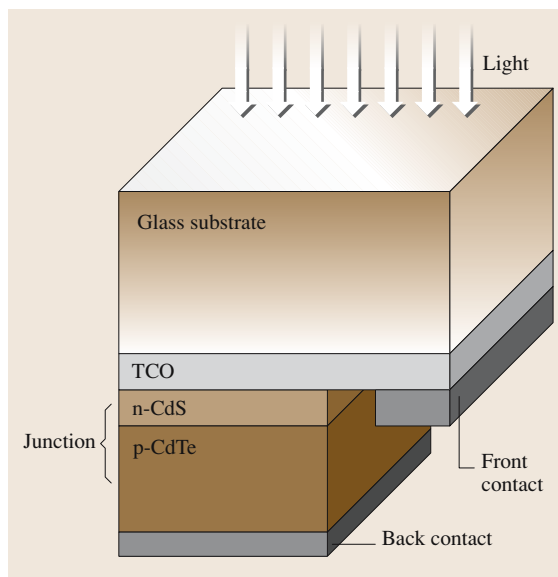


Fig. 46.8 Schematic of a CdTe device structure based on a glass superstrate

most common method for depositing CdS is by chemical bath deposition (CBD). The layer is n-type, forming a junction with the p-type CdTe. The thickness is kept to a minimum (around 100–200 nm) in order to minimise absorption at the blue end of the spectrum. A schematic of the CdTe PV structure is shown in Fig. 46.9.

Deposition techniques for CdTe (electrodeposition or close-space sublimation (CSS)) require post-deposition treatment for doping, grain growth and resistivity reduction, at temperatures above 400 °C. The exact processes are not clear but improve the efficiency from a few % to > 10% if this anneal is carried out in air in the presence of CdCl₂ [46.10]. There are variations in this process in terms of temperature, time and ambient gas but all involve the diffusion of Cl into the CdTe and some interdiffusion of the CdS/CdTe interface. This process improves grain size to $\approx 1\text{ }\mu\text{m}$ if the starting material has a much smaller grain size. This is further complicated by the variations in grain size from the CdS interface up to the back surface, where the junction region displays smaller grains. The best efficiency in the laboratory is 16.5% reported by Wu [46.11]. Production-scale processes achieve somewhat lower efficiencies, typically 8–10% [46.12].

A typical CdTe module with an area of 0.94 m² has a peak output power of 70 W; this compares favourably with other thin-film photovoltaics but is still much lower than the crystalline silicon modules [46.13]. However, the cost advantage and long-term stability make CdTe modules an attractive candidate for terrestrial applications. Future prospects for CdTe can be seen in current research activities with encouraging results for CdTe solar cells on flexible plastic substrates [46.14]. Using these substrates places tighter constraints on deposition and annealing temperatures and alternatives to the CdCl₂ anneal would be attractive. One prospect is for in situ doping of CdTe with arsenic to make it p-type using metalorganic chemical vapour deposition (MOCVD). This technique has been successfully used in the past to deposit films that were given a conventional anneal treatment. MOCVD offers greater flexibility over deposition conditions but would also need to compete with CSS and electrodeposition on cost of deposition.

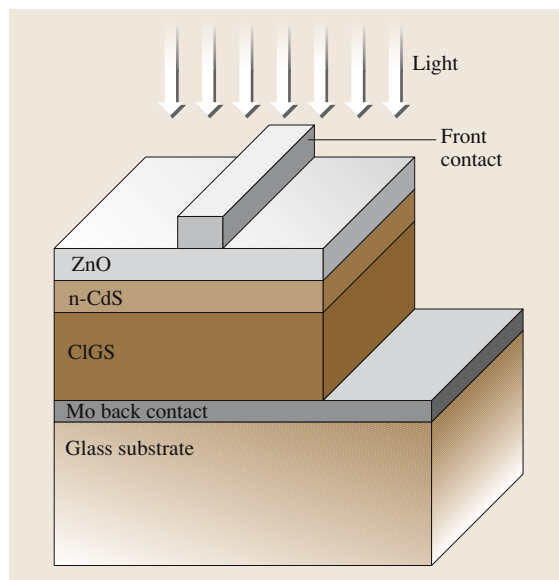


Fig. 46.9 Schematic of CIGS device structure based on a glass substrate

The abundance figures look low but up to 5000 tons of Te could be produced each year from copper ore and 20 000 tons of Cd is produced each year as a by-product of zinc production. An important environmental and resource issue may be the recycling of these materials. It is important to note that although Cd is a very toxic metal it is relatively benign in the form of CdTe as it is a very stable compound and is not soluble in water. The use of Cd in solar cells has the added benefit of making use of the Cd that is produced as a by-product of zinc production in an environmentally friendly way. Jager-Waldau [46.6] states that “Every energy source or product may present some environmental, health and safety hazard, and those of CdTe should by no means be considered barriers to technology scaling”. This technology is less well developed than those in the other case studies but the thin-film technology on glass combined with good stability makes this a very attractive proposition for both terrestrial and space applications.

46.6 CuInGaSe₂ (CIGS) Thin-Film Solar Cells

This is an alternative material system for achieving an optimum band gap for solar energy conversion and one that has attracted a lot of research effort in the

past 10 years with impressive improvements in conversion efficiency. The highest efficiency reported for laboratory solar cells is 19.2%, which is currently the

best of the thin-film PV technologies [46.15]. In common with other thin-film PV technologies, these films can be deposited onto cheap substrates, at relatively low temperature, and the potential for processing in large volumes. The absorber layer is based on the Chalcopyrite-phase CuInSe_2 (CIS). This has a band gap of 1.04 eV, which is lower than the optimum band gap for optimum efficiency. However, cells made from CIS have achieved > 10% efficiency. Alloying with Ga to form $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ (CIGS) increases the band gap to 1.7 eV for $x = 1$. This gives a range of alloy composition across the important, optimum, range for efficient conversion. An increase in band gap will increase the voltage of the cell (V_{oc}) but decrease the number of absorbed photons, thus decreasing J_{sc} . In practice the alloy is not uniform and thus a greater proportion of the photon flux is absorbed by the single-junction device than would occur for a single wide-band-gap absorber layer [46.16]. The other benefit of grading of the alloy composition is that the resultant band-gap grading creates a built-in electric field that can drift the electrons towards the junction.

Unlike a-Si and CdTe, the preferred arrangement is not to illuminate through the substrate but illuminate from the top surface as with the crystalline GaAs cells. The typical layer structure is shown in Fig. 46.9. The substrate is soda-lime glass with a sputtered coating of molybdenum which acts as the back contact. The next layer is the $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ alloy, which is the p-type absorber layer. The junction is formed with a thin layer of CdS (as with CdTe). The front contact is formed by aluminium-doped ZnO, which is highly conducting but transparent, allowing solar radiation to pass without significant absorption. One disadvantage of the substrate approach is that the top surface is protected with another glass sheet when it is fabricated into a solar module. This step is not necessary in the superstrate approach. However, the substrate configuration has shown much higher maximum efficiencies than the superstrate approach and offers the opportunity to deposit the CIGS layers onto flexible metal foils. The structure for CIGS on metal foil substrates is similar to those on the Mo/glass substrates except for a barrier layer of SiO_2 between the metal foil and the Mo back contact. This is needed to reduce the diffusion of impurities into the active semiconductor

layer. However, one impurity, sodium, has to be put back in to give the desired doping properties for the CIGS absorber layer [46.17]. This impurity occurs naturally when cheap soda-lime substrates are used, as the sodium will diffuse from the glass substrate into the CIGS layer during film deposition and annealing. Using metal foil substrates makes the PV cells flexible and enables their manufacture in a continuous reel-to-reel process.

There are various deposition methods that can be used for the deposition of CIGS and this can lead to low-cost production routes. The early results were obtained by co-evaporation from separate elemental sources. More recently a range of techniques have been used from e-beam evaporation of the metal sources and subsequent selenisation and annealing to electrodeposition. The characteristic of these approaches is a separation of the deposition process and alloy formation, allowing cheap and potentially high-throughput techniques to be used. This also enables control of the alloy composition and stoichiometry that will affect the doping [46.18]. The subsequently deposited junction layer, CdS, can be deposited by either chemical bath deposition, sputtering or CVD. The transparent contact layer, ZnO, is typically deposited by sputtering. All these techniques can be operated on a large industrial scale and lend themselves to cheap production methodology.

CIGS solar cells have moved rapidly into production and a number of manufacturers are offering modules with efficiencies in excess of 10%. For example, Shell Solar produce CIGS panel with a peak output power of 40 W for a module area of 0.42 m^2 . The potential price reduction is similar to other thin-film technologies and with a potential for breaking the critical 1 \$/W barrier, along with a-Si and CdTe. The current production cost is three to four times this value.

As with CdTe solar cells, the abundance of the elements In and Se appears low, but for a thin-film photovoltaic, where each module only needs a few grams of material, the total amount needed is not huge. However, the current price of In is high as it is not readily produced as a by-product of other mining processes. The cell stability for terrestrial applications appears to be good and has potential for space applications but more needs to be understood about the effects of high-energy γ fluxes on long-term stability.

46.7 Conclusions

This chapter has introduced some basic concepts about photovoltaic solar cells and examples of some of the

more common materials being used for solar-cell production. The area of research is huge with a number

of specialist journals dedicated to solar energy and large international conferences held annually. The drive to find cheaper solutions to the conversion of solar energy has taken research down some interesting and unusual avenues. One of the more successful of the alternative approaches is the dye-sensitised cell, known as the Gratzel cell after its inventor *Michael Gratzel* [46.19]. This uses organic dyes to absorb the solar radiation and transfer electrons to a porous TiO_2 surface, which is effectively the junction. Other approaches have looked at conducting polymers and polymer blends to form photovoltaic junctions. These approaches could enable very cheap photovoltaic solar cells to become a reality but over a much longer timescale than for the materials considered in this chapter.

All the materials considered here are in production and look set to make a major contribution to the production of solar energy over the next 10 years. The largest and most mature production facilities are based on crystalline or multicrystalline silicon. These modules are also the most efficient for single-junction cells. The most efficient modules are made from GaAs on Ge and are the most complex and most expensive. Most of this production is for the space market but increasingly is showing potential for use with concentrators for terrestrial appli-

cations. This is acceptable for stand-alone solar power generation but the largest challenge for large-scale terrestrial power generation is to integrate photovoltaic modules into buildings. The equivalent of a gigawatt power station would require, at least, of the order of 10^7 m^2 of solar modules. In reality it would have to be much larger to account for the average power production being much less than the peak production. A modest installation would have approximately 100 m^2 of solar panels so we would need more than 10^5 buildings. So, the goal has to be that every house and building has a PV façade or roof.

Large-scale implementation will require the modules to be architecturally acceptable, in appearance and size, and probably serve multiple functions such as keeping the rain out and heat in etc. This need is opening the opportunity for a range of materials and designs and ultimately flexible panels. On top of this the cost per watt of electricity produced needs to come down, probably by a factor of five based on current installed costs. The thin-film technologies are inherently cheaper but much of this cost benefit is not realised at present because the efficiencies and production volumes are low. The current growth in PV solar energy after decades of research has become very rapid and this expansion is set to continue well into the future.

References

- 46.1 C. H. Henry: *J. Appl. Phys.* **51**, 4494 (1980)
- 46.2 J. Zhao, A. Wang, M. A. Green: *Appl. Phys. Lett.* **73**, 1991 (1998)
- 46.3 L. Pirozzi, G. Arabito, F. Artuso, V. Barbarossa, U. Besi-Vetrella, S. Loreti, P. Mangiapane, E. Salza: *Sol. Energy Mater. Sol. Cells* **65**, 287 (2001)
- 46.4 A. G. Aberle: *Sol. Energy Mater. Sol. Cells* **65**, 239 (2001)
- 46.5 N. Bernhard, G. H. Bauer, W. H. Bloss: *Prog. Photovolt. Res. Appl.* **3**, 3 (1995)
- 46.6 A. Jager-Waldau: *Sol. Energy* **77**, 667 (2004)
- 46.7 N. H. Karam, R. R. King, M. Haddad, J. H. Ermer, H. Yoon, H. L. Cotal, R. Sudharsanan, J. W. Eldredge, K. Edmondson, D. E. Joslin, D. D. Krut, M. Takahashi, W. Nishikawa, M. Gillanders, J. Granata, P. Hebert, B. T. Cavicchi, D. R. Lillington: *Sol. Energy Mater. Sol. Cells* **66**, 453 (2001)
- 46.8 L. M. Fraas, J. E. Avery: *Optoelectron. Dev. Technol.* **5**, 297 (1990)
- 46.9 J. F. Geisz, D. J. Friedman, J. M. Olson, S. R. Kurtz, B. M. Keyes: *J. Cryst. Growth* **195**, 401 (1998)
- 46.10 X. Wu, J. C. Keane, R. G. Dhere, C. Dehart, D. S. Albin, A. Dude, T. A. Gessert, S. Asher, D. H. Levi, P. Sheldon: *Proc. 17th Eur. Photovolt. Sol. Energy Conf.*, **II**, 995 (2001)
- 46.11 X. Wu: *Sol. Energy* **77**, 803 (2004)
- 46.12 A. K. Turner, J. M. Woodcock, M. E. Ozsan, D. W. Cunningham, D. R. Johnson, R. J. Marshall, N. B. Mason, S. Oktik, M. H. Patterson, S. J. Ransome, S. Roberts, M. Sadeghi, J. M. Sherborne, D. Sivapathasundaram: *Sol. Energy Mater. Sol. Cells* **35**, 263 (1994)
- 46.13 D. Cunningham, K. Davies, I. Grimmond, J. Healy, E. Mopas, N. O'Connor, M. Rubeich, M. Sadeghi, D. Skinner, T. Trumbly: *16th Eur. Photovolt. Sol. Energy Conf.*, 1–5 May, Proceeding of the sixteenth European Photovoltaic Solar Energy Conference, Glasgow, UK 2000, ed. by H. Scheer, B. McNelis, H. A. Ossenbrink, P. Helm (James & James (Science Publishers) LTD., Glasgow 2000) 281
- 46.14 X. Mathew, J. P. Enriquez, A. Romeo, A. N. Tiwari: *Sol. Energy* **77**, 831 (2004)
- 46.15 K. Ramanathan, M. A. Contreras, C. L. Perkins, S. Asher, F. S. Hasoon, J. Keane, D. Young, M. Romero, W. Metzger, R. Noufi, J. Ward, A. Duda: *Prog. Photovolt.: Res. Appl.* **11**, 225 (2003)

- 46.16 I. M. Kotschau, G. Bilger, H. W. Schock: Mater. Res. Soc. Symp. Proc. **763**, 263 (2003)
- 46.17 D. Hermann, F. Kessler, K. Hertz, M. Powalla, A. Schulz, J. Schneider, U. Schumacher: Mater. Res. Soc. Symp. Proc. **763**, 287 (2003)
- 46.18 S. Hishikawa, T. Satoh, S. Hayashi, Y. Hashimoto, S. Shimakawa, T. Megami, T. Wada: Sol. Energy Mater. Sol. Cells **67**, 217 (2001)
- 46.19 M. Gratzel: MRS Bull. **30**, 23 (2005)

47. Silicon on Mechanically Flexible Substrates for Large-Area Electronics

Low-temperature thin-film semiconductors and dielectrics are critical requirements for large-area electronics, including displays and imagers. Despite the presence of structural disorder, these materials show promising electronic transport properties that are vital for devices such as thin-film transistors (TFTs). This chapter presents an overview of material and transport properties pertinent to large-area electronics on mechanically flexible substrates. We begin with a summary of process challenges for low-temperature fabrication of a-Si:H TFTs on plastic substrates, followed by a description of transport properties of amorphous semiconducting films, along with their influence on TFT characteristics. The TFTs must maintain electrical integrity under mechanical stress, induced by bending of the

47.1	a-Si:H TFTs on Flexible Substrates	1108
47.2	Field-Effect Transport in Amorphous Films	1108
47.2.1	Localized and Extended States	1109
47.2.2	Density of States (DOS)	1110
47.2.3	Effective Carrier Mobility	1110
47.3	Electronic Transport Under Mechanical Stress	1113
47.3.1	Thin-Film Strain Gauges	1114
47.3.2	Strained Amorphous-Silicon Transistors	1116
	References	1118

flexible substrates. Bending-induced changes are not limited to alteration of device dimensions and involve modulation of electronic transport of the active semiconducting layer.

There is growing interest in low-temperature thin-film semiconductors and dielectrics for a variety of applications such as organic light-emitting diode (OLED) displays and lighting modules [47.1, 2], solar cells [47.3], and digital imagers [47.4, 5]. Interest in these materials is driven by the promise of low-cost roll-to-roll manufacturing [47.6]. They enable high-performance electronic devices including thin-film transistors (TFTs), OLEDs, and p-i-n photodetectors on large-area mechanically flexible plastic substrates. OLEDs are becoming increasingly attractive as emission devices due to their fast response, high conversion efficiency, wide viewing angle, and compatibility with plastic substrates [47.7]. In high-information-content displays, the OLED must be integrated with a TFT circuit that provides stable drive currents [47.8]. Here, low-temperature a-Si:H TFTs or (solution-processed/vacuum-deposited) organic TFTs can be used to realize these circuits.

In addition to advances in material deposition and integration technologies, significant research efforts are being undertaken to improve the lifetime of the TFT and the OLED in view of instabilities induced by bias stress and exposure to ambient con-

ditions. The instability of TFTs stems from structural and interfacial disorder in the semiconductor and insulator layers, and interfaces. This leads to a shift in the threshold voltage ΔV_T under prolonged bias stress in a-Si:H [47.9] and organic [47.10] TFTs. To deal with the shift in device parameters, pixel circuits and driving schemes that can provide compensation have been proposed and are well under investigation [47.11]. The lifetime of the OLED, on the other hand, can be enhanced by efficient thin-film encapsulation layers with water-vapor permeation rates $< 10^{-6}$ g/m²/day at 25 °C and less than 40% relative humidity [47.12], which are imperative for active matrix organic light-emitting diode (AMOLED) displays on plastic substrates.

This chapter presents an overview of material and transport properties pertinent to mechanically flexible amorphous-silicon electronics. We begin with an overview of low-temperature fabrication of high-performance a-Si:H transistors and associated processing challenges. The transport properties of amorphous semiconducting films are then presented, along with their influence on TFT characteristics. We also examine the impact of external mechanical stress,

induced by bending of the flexible substrate, on the characteristics of thin-film devices, including strain gauges and TFTs. The shifts are not only limited to

changes in device dimensions but also to a modulation of electronic transport of the active semiconducting layer.

47.1 a-Si:H TFTs on Flexible Substrates

Compared to conventional (rigid) glass substrates, mechanically flexible substrates are attractive because of their reduced fragility and weight [47.13]. Polyethylene terephthalate (PET) [47.14] and polyimide [47.15] are examples of such substrates. Most plastic substrates have a low glass-transition temperature T_g ($< 250^\circ\text{C}$). Consequently, the device processing temperatures must be reduced without compromising the electronic properties of the deposited materials. Despite the adverse effects of reducing the processing temperature, high-performance low-temperature ($< 150^\circ\text{C}$) processes have been developed for fabrication of a-Si:H TFTs on flexible substrates [47.14–16], and even lower temperature processes (75°C) are under investigation [47.17].

One of the key challenges in low-temperature fabrication is to achieve a high-quality a-SiO_x [47.17] or a-SiN_x:H [47.16] gate and/or passivation dielectric with low defect density, leakage, and interface state density. Nitrogen rich ($x = 1.6$ to 1.7) a-SiN_x:H films with high resistivity ($\rho > 10^{15} \Omega\text{-cm}$) and breakdown field ($> 5 \text{ MV/cm}$) have been deposited using conventional 13.56 MHz plasma-enhanced chemical vapor deposition (PECVD) at 120°C [47.16].

The fabrication challenges are not limited to the electrical quality of the deposited materials. Most plastic substrates have a large coefficient of thermal expansion (CTE) ($> 50 \text{ ppm}/^\circ\text{C}$), which is much higher than that of the TFT layers ($\approx 4 \text{ ppm}/^\circ\text{C}$) [47.18]. As the deposited films go through a temperature drop ΔT upon cooling from the process temperature [47.2], thermal

stress $\sigma_{\text{th}} = Y_f \Delta C \Delta T$ develops in the film [47.19]. Here, ΔC is the difference between CTEs of the substrate and the film, and Y_f denotes the biaxial elastic modulus of the deposited film. The thermal stress bends the composite structure, giving rise to handling problems. Thus, the substrate needs to be held flat in a frame or glued to a rigid substrate [47.2]. In addition, since most layers are patterned during processing, the thermal stress exhibits a local pattern-dependent distribution, which leads to problems in alignment of consecutive masks. Lowering the process temperature and choosing a substrate with a smaller CTE, e.g. polyethylene naphthalate (PEN) with a CTE of $13 \text{ ppm}/^\circ\text{C}$ [47.18], will help alleviate the thermal stress, and associated alignment issues. It is important to note that thin rigid (steel, aluminum, or glass) substrates are also mechanically flexible and are being considered for device fabrication [47.20]. Although these substrates can withstand higher temperatures, handling issues still remain and depend on the desired degree of mechanical flexibility.

Recent work on TFT fabrication on flexible substrates is starting to demonstrate successful handling of material processing issues. For instance, TFTs fabricated on plastic substrate at 150°C have been reported to have parameter values very similar to the high-temperature counterparts in terms of low reverse and high on current (with switching ratio of $> 10^8$), high field-effect mobility $> 0.8 \text{ cm}^2/\text{Vs}$, threshold voltage of $\approx 1\text{--}2 \text{ V}$, and low subthreshold slope $< 0.3 \text{ V/dec}$ [47.16].

47.2 Field-Effect Transport in Amorphous Films

The electronic properties of a semiconducting film are primarily determined by its structural disorder and grain size, which strongly depend on deposition conditions and postprocessing treatment [47.21].

The structure of disordered semiconductors such as a-Si:H lacks the long-range periodicity of crystalline silicon. However, the atoms of the materials bond to their neighbors according to the 8-N rule [47.22]

for both amorphous and crystalline silicon, leading to short-range structural order. The extent to which this short-range order is conserved is dependent on processing conditions and leads to different degrees of disorder. This range can be as small as a few nanometers in amorphous semiconductors or as large as a few tens of microns in polycrystalline material. Nanocrystalline (nc) and microcrystalline (μc) materials reside

between these extremes depending on grain size and stoichiometry.

The observation of an absorption gap and thermally activated conductivity in amorphous semiconductors stems from the short-range order [47.24] and resembles the band-gap properties of an ideal crystalline semiconductor, which has no states in the gap. The edges of the gap are well defined and represent discontinuities in the density of states (DOS). In contrast, the lack of long-range order in the amorphous semiconductor leads to the presence of states in the gap, and thus, tailing of the band edges.

47.2.1 Localized and Extended States

In the ideal crystalline semiconductor, the band states are all extended with a corresponding band mobility μ_{band} . In the case of an amorphous semiconductor, however, the structural disorder of the material leads to the presence of localized states for which the mobil-

ity vanishes at $T = 0$. Anderson [47.25] demonstrated that, for a sufficiently large degree of disorder, all states in the band are localized. In his work, the disorder in the material is modeled by a random but uniform distribution U_0 for the energy of the wells at lattice sites. He found that all of the states of the energy band become localized when $U_0/B > 5$, where B is the band width and is given by $B = 2ZI$ [47.22]. Here, Z is the coordination number, and I the overlap energy for the states in neighboring wells, which decreases exponentially with distance based on the tight-binding approximation.

For a small degree of disorder, Anderson's criteria may not hold for states with large overlap energies (i. e. the extended states). Mott and Davis [47.22] have stipulated that, for a given value of U_0 , the states can be localized or extended, depending on their overlap energy. Consequently, for some materials there are distinct energies which separate the localized states from the extended ones and can be determined from Anderson's criteria $U_0 \approx 5B$. These energies are referred to as mobility edges (or mobility shoulders) due to the significant change in the mobility expected at these edges. Figure 47.1 illustrates the DOS $g(E)$ for the mobility gap of a-Si:H indicating the extended and localized states. The mobility shoulder E_C of the conduction band separates the localized states from the extended ones. A similar band structure and mobility shoulder is present for the valence band. This graph is compiled from the work of Cohen et al. [47.23], who first demonstrated that the change in mobility at these edges can be of the order of 10^3 – 10^4 , signifying the different transport mechanisms on the two sides of the edge.

In the extended states, the carriers have a finite mobility, and thus take part in metallic-type conduction. This prevails at room temperature for a-Si:H and takes the form $\sigma_0 = 0.06q^2/\hbar r(E)$ [47.22], where q is the elementary charge and $r(E)$ the mean distance between the wells that contribute to states with energy E . In the conduction band, $r(E)$ will be close to the lattice constant, which yields $\sigma_0 = 350 \Omega^{-1} \text{cm}^{-1}$. For Fermi energies $E_F < E_C$, the conductivity depends on the Fermi energy and reads

$$\sigma = \sigma_0 \exp(E_F/k_B T), \quad (47.1)$$

where k_B is Boltzmann's constant and T the temperature, predicting Arrhenius behavior for the conductivity. Here, E_F is defined as negative for the traps with respect to the conduction band edge E_C . This change in conductivity is due to a change in the concentration of carriers, excited to the extended states n_{band} with a band mobility

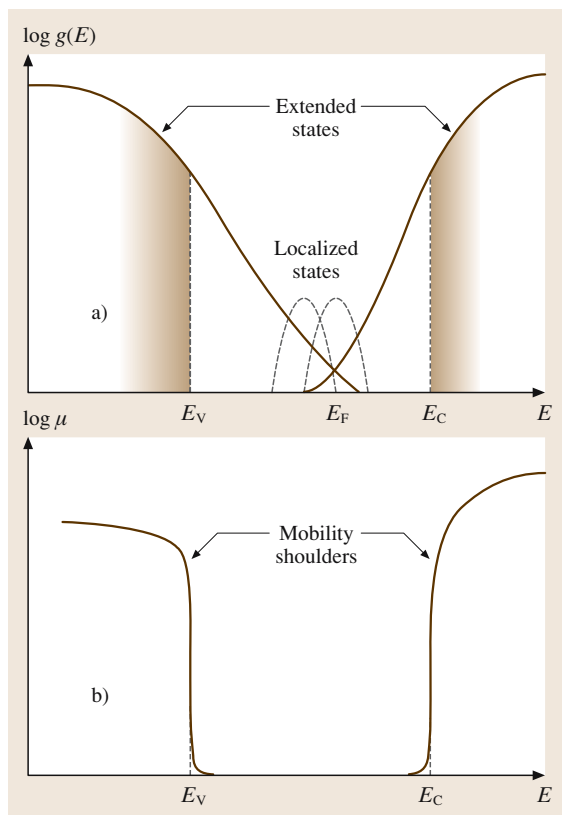


Fig. 47.1a,b Sketch of (a) DOS in valence and conduction bands, and (b) hole and electron mobilities (after [47.23])

μ_{band} such that

$$\sigma = q\mu_{\text{band}}n_{\text{band}}, \quad (47.2)$$

where

$$n_{\text{band}} = N_b \exp(E_F/k_B T), \quad (47.3)$$

with $N_b = \sigma_0/q\mu_{\text{band}}$. From the given value for σ_0 , the band mobility can be found as $\mu_{\text{band}} \approx 12 \text{ cm}^2/\text{Vs}$ for a-Si:H.

In the localized states, the conduction is only possible by hopping of the carriers to the neighboring localized states [47.26]. This method of conduction prevails at low temperatures ($T < 200 \text{ K}$) and for highly disordered materials.

The picture of the mobility gap can be completed by adding defect states in the middle of the gap, as indicated by the dashed lines in Fig. 47.1a. These states are attributed to the broken or unsaturated bonds such as dangling bonds in the network structure. The density of dangling bonds is reduced to $10^{15} - 10^{16} \text{ cm}^{-3}$ in a-Si:H by hydrogen atoms, which passivate some of these unsaturated bonds [47.24, 27].

47.2.2 Density of States (DOS)

The density of states (DOS) $g(E)$, including the density of band tail and defect states, as schematically represented in Fig. 47.1a, determines the transport properties of the disordered semiconductor. The DOS in

a-Si:H has been extensively studied by using different experimental techniques such as field-effect measurements [47.28, 29], photoconductivity measurements [47.30], deep-level transient capacitance spectroscopy (DLTS) [47.31], and capacitance–voltage–frequency ($C-V-f$) characteristics [47.32, 33]. It is found that an exponential distribution of the deep and tail states [47.34] can efficiently describe the distribution of the localized states and hold for a wide range of materials with rapidly changing DOS. In the case of a-Si:H, the localized states in the upper half of the mobility gap, closer to E_C , behave as acceptor-like states, while the states in the lower part of the gap, closer to E_V , behave as donor-like states. Acceptor-like states are neutral when they are empty, and negatively charged after capturing an electron, whereas the donor-like states are positively charged when they are empty and neutral after capturing an electron.

In a-Si:H, the number of donor-like states closer to the valence band is much higher than the number of acceptor-like states. As a result, following the neutrality condition, the position of the Fermi energy E_i in an intrinsic a-Si:H sample in the dark lies closer to E_C due to the asymmetrical DOS distribution [47.27]. This results in a much stronger electron conduction in a-Si:H, which signifies the role of the conduction band tail in dispersive electron transport. The density of states for the conduction band tail $g(E)$ is written as

$$g(E) = N_t/k_B T_t \exp(E/k_B T_t), \quad (47.4)$$

where N_t is the total acceptor-like states in the conduction band tail, and T_t the associated slope of the exponential state distribution.

47.2.3 Effective Carrier Mobility

Due to the high density of tail and deep states, only a small number of carriers are thermally excited to the extended states and contribute to conduction as described in (47.2). The total carrier density n is the sum of the excited and trapped carriers such that

$$n = n_{\text{band}} + n_t. \quad (47.5)$$

To obtain the trapped carrier density n_t as a function of the Fermi energy E_F , we integrate the product of the DOS and the probability of occupation of a state $f(E)$ over the mobility gap

$$n_t = \int_{E_V}^0 g(E)f(E)dE. \quad (47.6)$$

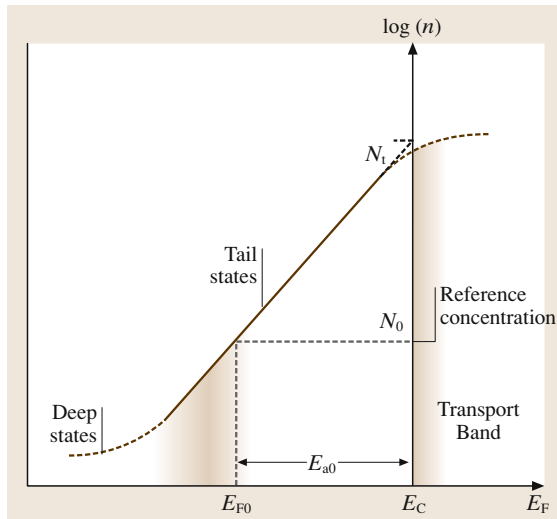


Fig. 47.2 Density of trapped carriers as a function of the Fermi energy, showing the reference concentration N_0 and transport (conduction) band

In equilibrium, the probability of occupation can be described by the Fermi–Dirac function,

$$f(E) = \frac{1}{1 + \exp[(E - E_F)/k_B T]} . \quad (47.7)$$

Equation (47.6) can be numerically solved and approximated along the lines given by *Shaw and Hack* [47.37] as

$$n_t(E_F) = N_t \exp(E_F/k_B T_i) u(T_i/T, E_F) . \quad (47.8)$$

The underlying assumption is that E_F moves no closer than a few $k_B T$ to the mobility edge. This is true because of the high density of tail states, which tends to pin the movement of the Fermi energy. Here, $u(T_i/T, E_F)$ represents the changes in the trapped-carrier density with normalized ambient temperature T_i/T . For $T \ll T_i$, $u(T_i/T, E_F)$ is often approximated by $[\sin(\pi T/T_i)/\pi T/T_i]$ [47.38] with a value close to 1. The ratio T_i/T , referred to as the dispersion parameter, characterizes the dispersive transport of electrons in the conduction band tail, and can be obtained from the time dependence of the electron drift mobility in time-of-flight experiments [47.39]. The presence of a high trapped-carrier density in amorphous semiconductors leads to an effective trapped-carrier mobility that is lower than the band mobility. *Street* [47.27] has defined the drift mobility μ_D of the carriers as the band mobility reduced by the fraction of time that the carrier is trapped,

$$\mu_D = \mu_{\text{band}} \frac{\tau_{\text{band}}}{\tau_{\text{band}} + \tau_{\text{trap}}} , \quad (47.9)$$

where τ_{band} and τ_{trap} are the times that carriers spend in the extended and localized states, respectively.

Carriers accumulated by the field effect in TFTs also demonstrate similar transport properties. The field-effect mobility μ_{FE} in TFTs is conventionally defined as

$$\mu_{\text{FE}} = g/qn , \quad (47.10)$$

where g is the conductance of the film, and qn is the field-induced charge in the semiconducting film. The terms g and n are averaged over the device active area and are related to the number of band carriers n_{band} and trapped carriers n_t . The μ_{FE} is conventionally retrieved from measurement of the transistor current ($I_{\text{DS,lin}}$) in the linear regime ($V_{\text{DS}} = 0.1$ V), by using the following:

$$\mu_{\text{FE}} = \frac{L}{WC_i V_{\text{DS}}} \frac{\partial I_{\text{DS,lin}}}{\partial V_{\text{GS}}} , \quad (47.11)$$

where W and L are the channel width and length, respectively, C_i the gate capacitance, and V_{GS} and V_{DS} are the gate–source and drain–source biases, respectively.

According to (47.3) and (47.8), the densities of the band and trapped carriers increase differently with increasing E_F . Consequently, when n_{band} and n_t are averaged over the volume of the semiconducting film to obtain the μ_{FE} described by (47.10), the mobility becomes a function of device parameters (e.g. layer thicknesses and bias conditions).

The bias dependence is also evident in temperature-dependence measurements of mobility and conductivity in a-Si:H and organic TFTs. Figure 47.3 illustrates the temperature dependence of μ_{FE} for a-Si:H, pentacene and polythienylene vinylene (PTV) [47.36], and poly[5,5'-bis(3-alkyl-2-thienyl)-2,2'-bithiophene] (PQT-12) [47.35,40] TFTs at different gate biases. Here, the activation energy E_a turns out to be bias-dependent (inset of Fig. 47.3). More importantly, an anomaly arises in which the μ_{FE} and E_a become dependent on the gate capacitance C_i , which is solely a geometrical capaci-

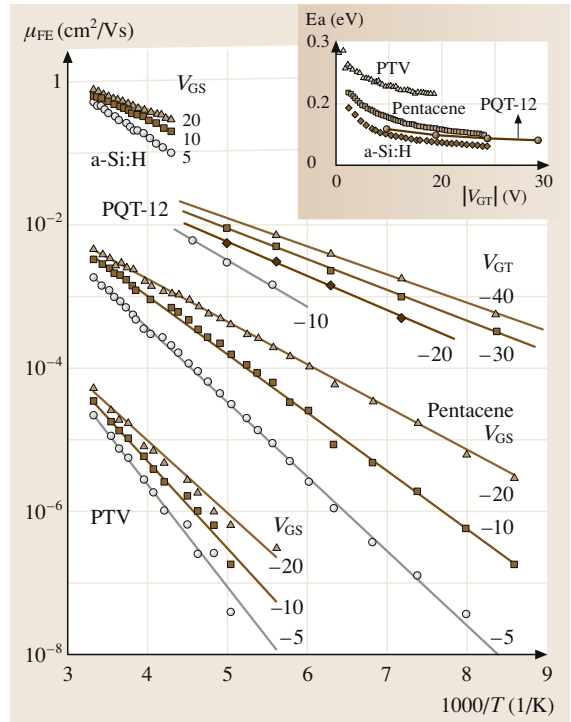


Fig. 47.3 Temperature dependence of μ_{FE} for a-Si:H, PQT-12 [47.35], pentacene and PTV [47.36]. Inset shows the bias dependence of the activation energy E_a of μ_{FE} in these materials

tance. A higher C_i implies higher carrier accumulation and consequently higher μ_{FE} for the same bias.

To remove this anomaly, we need to identify the effective carrier mobility such that $\mu_{FE} = \mu_{eff} \times f(\phi)$, where μ_{eff} is the effective physical mobility and $f(\phi)$ describes the device attributes such as bias and geometry.

We reconsider the densities of trapped and free carriers according to (47.3) and (47.8). Since both densities vary exponentially with the Fermi energy, the density of band carriers in terms of trapped carriers can be written as

$$n_{band} = \theta n_t^{T_i/T}, \quad (47.12)$$

where $\theta = N_b/(N_t u)^{T_i/T}$. Equation (47.12) has been found to hold empirically for a wide range of disordered semiconductor systems and operating conditions, including organic semiconductors [47.41, 42].

According to (47.2) and (47.12), the conductivity $\sigma(n)$ as a function of carrier concentration can be written as

$$\sigma(n) = q\mu_{eff} N_0 (n/N_0)^{T_i/T}, \quad (47.13)$$

where μ_{eff} is the effective mobility defined at a reference concentration as

$$\mu_{eff} \equiv \frac{\sigma(n = N_0)}{qN_0} = \mu_{band} \frac{N_b}{N_0} \left(\frac{N_0}{N_t u} \right)^{T_i/T}. \quad (47.14)$$

Consequently, the conductivity at any carrier concentration is given by μ_{eff} and T_i/T . Here, μ_{eff} represents the effective carrier mobility at concentration N_0 and T_i/T describes the change in conductivity with carrier concentration.

The significance of this representation becomes clear when we use it to obtain the current–voltage characteristics of a TFT (Fig. 47.4). Using the gradual channel

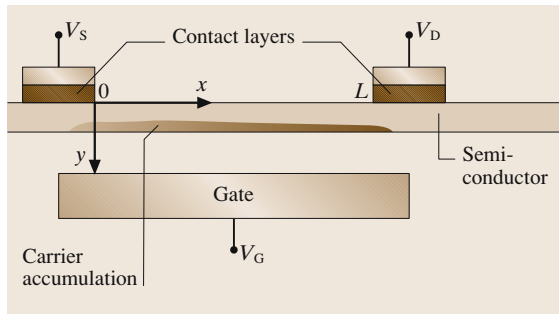


Fig. 47.4 Simplified schematic of an inverted staggered TFT

approximation, we can write

$$I_{DS,lin} = \frac{W}{L} V_{DS} \int_0^\delta \sigma(y) dy, \quad (47.15)$$

where y denotes the location across the channel and δ the channel depth. Using (47.13) for σ and changing the integral parameter from y to n , we find after mathematical manipulation [47.43]

$$I_{DS,lin} = \mu_{eff} \zeta \frac{W}{L} (C_i V_{GT})^{\alpha-1} V_{DS}, \quad (47.16)$$

where $V_{GT} = V_{GS} - V_{thr}$, V_{thr} the threshold voltage, $\alpha = 2T_i/T$ the saturation current–voltage characteristics power parameter, and

$$\zeta = \frac{(2\epsilon k_B T_i N_0)^{1-\alpha/2}}{\alpha - 1}. \quad (47.17)$$

Here, ζ is just a function of T_i/T and accounts for the carrier distribution across the film. Revisiting the definition of the field-effect mobility as given by (47.11), but this time employing a more meaningful representation of the current, i. e. (47.16), we find

$$f(\phi) \equiv \frac{\mu_{FE}}{\mu_{eff}} = \zeta (\alpha - 1) (C_i V_{GT})^{\alpha-2}. \quad (47.18)$$

Representation of μ_{eff} according to (47.14) is valid only when there is an exponential relationship between the carrier concentration n and the Fermi energy E_F as given by (47.8). This is true for Fermi energy locations below the transport band edge and above the energies of the deep states (the region shown by the solid slope in Fig. 47.2). Since the deep states mostly contribute to the threshold voltage V_{thr} and are filled before the device turns on, the mobility definition of (47.14) is valid for the above-threshold regime. Consequently, the value of the reference concentration N_0 must be selected such that the Fermi energy E_{F0} associated with N_0 resides well above the deep states. This requires the charge accumulated in the channel $Q_{channel}$ to be higher than the charge $C_i V_{thr}$ needed to turn on the device. If the carrier concentration at the semiconductor interface is N_0 , $Q_{channel} = C_i V_0$, where $V_0 = (2\epsilon k_B T_i N_0)^{1/2}/C_i$. Thus, for $Q_{channel} > C_i V_{thr}$, we conclude $V_0 > V_{thr}$ or $N_0 > C_i^2 V_{thr}^2 / (2\epsilon k_B T_i)$, indicating the lower limit for N_0 . Using typical values, $V_{thr} = 2$ V and $C_i = 20$ nF/cm², we have $N_0 > 6 \times 10^{16}$ cm⁻³. For instance, with $N_0 = 10^{17}$ cm⁻³, we have for the bias dependence factor $\zeta = (4.1 \times 10^{-16} \alpha)^{(1-\alpha/2)} / (\alpha - 1)$.

We now revisit the temperature dependence of μ_{eff} by recasting (47.14) in the following manner

$$\mu_{eff} = \mu_{eff0} \exp(-E_{a0}/k_B T), \quad (47.19)$$

Table 47.1 Extracted transport parameters (μ_{eff} , T_i , and E_{a0}) for a selection of disordered semiconductors at $N_0 = 10^{17} \text{ cm}^{-3}$

Semiconductor	μ_{eff} at 300 K (cm^2/Vs)	T_i (K)	E_{a0} (eV)	Reference
a-Si:H	0.98	350	0.23	Our measurements
PQT-12	3.9×10^{-2}	326	0.24	[47.35]
Pentacene	9.0×10^{-4}	385	0.31	[47.36]
PTV	1.1×10^{-5}	380	0.42	[47.36]

where

$$\mu_{\text{eff}0} \equiv \mu_{\text{band}} \frac{N_b}{N_0} \quad \text{and} \quad E_{a0} \equiv k_B T_i \ln \left(\frac{N_i \mu}{N_0} \right) = -E_{F0}. \quad (47.20)$$

Equation (47.19) predicts Arrhenius behavior for the effective mobility with an activation energy of $E_{a0} = -E_{F0}$ (Fig. 47.2). The activation energy of μ_{eff} is not bias-dependent and corresponds to the energy needed for carriers to thermalize from E_{F0} to the mobility edge.

We now examine the relation between the two activation energies E_a and E_{a0} . To do so, we look at the temperature dependence of μ_{FE} and μ_{eff} , viz.,

$$\begin{aligned} \frac{\partial \mu_{\text{FE}}}{\mu_{\text{FE}} \partial T} &= \frac{\partial \mu_{\text{eff}}}{\mu_{\text{eff}} \partial T} + \frac{\partial f(\phi)}{f(\phi) \partial T} \\ &= -\frac{E_{a0}}{k_B T^2} + \frac{2T_i}{T^2} \ln \left| \frac{V_{\text{GT}}}{V_0} \right|, \end{aligned} \quad (47.21)$$

which yields

$$E_a = E_{a0} - 2k_B T_i \ln |V_{\text{GT}}/V_0|. \quad (47.22)$$

Equation (47.22) describes the bias dependence observed for E_a in the inset of Fig. 47.3.

Amorphous Organic Semiconductors

The concept of effective mobility and transport band may be generalized to accommodate different amorphous organic semiconductors despite differences in the underlying transport mechanism. This generalization follows from the relation $n_{\text{band}} = \theta n^{T_i/T}$ (47.12) for organic semiconductors for a wide range of temperatures and carrier concentrations [47.41]. For this relation

to hold, the distribution of trapped and band carriers must be exponential. Although evidence of a Gaussian trap distribution has been reported for organic materials [47.44], the Gaussian distribution is effectively seen as an equivalent exponential distribution due to the small variation in the Fermi energy because of the large tail state distribution. In addition, *Shapiro and Adler* [47.45] have demonstrated that a transport band is present in which hopping conduction dominates irrespective of the position of Fermi energy. Similar to the mobility edge, the trapped carriers are thermalized to the hopping band [47.26]. Relative to E_{F0} , the hopping band is located at

$$E_{0\text{hopping}} = k_B T_i \ln \left[\frac{\gamma^3}{N_0} \left(\frac{2T}{3T_i} \right)^3 \right], \quad (47.23)$$

where γ is the effective overlap parameter for electronic states in the band tail. *Baranovskii et al.* [47.46] have generalized this concept of transport band beyond the exponential DOS assumption and to a broader range of disordered materials with Gaussian or similar rapidly changing distributions. The generalized band concept can also accommodate the percolation-based hopping transport described by *Matters* [47.36] for amorphous organic semiconductors, which predicts a hopping band that is just $k_B T_i \ln (B_C/3\pi^3)$ higher than that predicted by *Monroe* [47.26]. Here, $B_C \approx 2.8$ is the critical number for percolation in three-dimensional amorphous systems. Table 47.1 summarizes the values for μ_{eff} at room temperature T_i , and E_{a0} at $N_0 = 10^{17} \text{ cm}^{-3}$, determined from the results presented for different disordered materials reported in literature.

47.3 Electronic Transport Under Mechanical Stress

Mechanical stress deforms the structure of the thin film leading to modulation in carrier mobility and density of states, and consequently, modulation of resistance.

The change in resistance of a solid with elastic strain or stress is commonly referred to as the *elastoresistance* or *piezoresistance* effect, respectively [47.47]. The

magnitude of the change is a function of the electronic properties of the material, the dimensions of the solid, and the direction of current flow.

Structural order in the material influences the elastoresistance effect. In the case of crystalline silicon, anisotropic scattering of electrons in the n-type material leads to a strong orientation dependence of the elastoresistive behavior [47.48]. In polycrystalline silicon, the crystallite size and orientation, and material texture play a critical role in determining the magnitude of the effect [47.49, 50]. In amorphous semiconductors, the random network behaves like an isotropic medium, and the anisotropy found in the crystalline material is less visible. However, the elastoresistance coefficients still depend on the relative orientation of the current and applied strain [47.51].

In sputtered amorphous silicon, *Welber and Brodsky* [47.52] have reported a decrease in the absorption gap with hydrostatic pressure with a coefficient of about -1 meV/kbar (-10^{-11} eV/Pa) observed from the shift in the absorption edge. *Weinstein* [47.53] has also reported similar results drawn from photoluminescence experiments. The change in the optical gap is similar in sign and magnitude to that measured for crystalline silicon [47.54]. *Lazarus* [47.55], however, has reported an exponential increase in the resistivity of a-Si:H with increasing hydrostatic pressure at room temperature. The increase in resistivity is ascribed to either a decrease in the number of carriers or a reduction of the mobility with compressive strain [47.51].

In this section, we investigate the impact of mechanical stress on electrical properties of thin-film devices, insight into which is critical for design of mechanically flexible electronics. We begin with metallic and semiconductor thin films for strain-gauge applications and continue with a-Si:H TFTs.

47.3.1 Thin-Film Strain Gauges

Strain gauges are thin-film transducer elements embedded on a substrate Fig. 47.5. The strain in the substrate leads to a change in the geometry, and therefore, the resistance of the gauge, which is detected and measured by external circuitry. The sensitivity of the resistance R to strain ϵ is referred to as the gauge factor k , which can be written as

$$k \equiv \frac{\Delta R}{\epsilon R} = \frac{\Delta L}{\epsilon L} - \frac{\Delta W}{\epsilon W} - \frac{\Delta t}{\epsilon t} + \frac{\Delta \rho}{\epsilon \rho}, \quad (47.24)$$

where ρ is the resistivity of the material, and W , L , and t are the width, length, and thickness of the strain

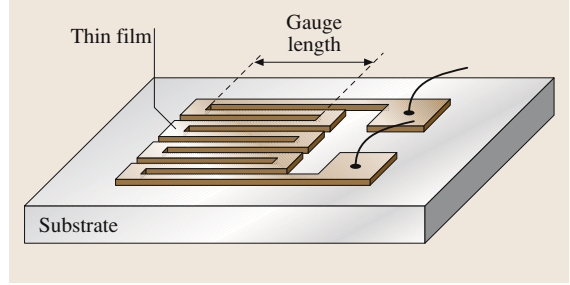


Fig. 47.5 Schematic of a thin-film strain gauge

gauge, respectively. The last term on the right-hand side of (47.24) reflects the strain-induced change in resistivity of the sample, whereas the first three terms refer to geometrical changes only.

Usually, the gauge length is oriented in a direction where strain is largest to achieve the highest gauge factor. This is referred to as the longitudinal orientation, where the current flows parallel to the strain. Similarly, the transverse orientation is defined when the gauge length is oriented in a direction perpendicular to the maximum strain (Fig. 47.6). The modulation of resistivity in the different orientations with the strain and correspondingly stress can be summarized by the following expression:

$$\frac{\Delta \rho_i}{\rho} = \gamma_{ij} \epsilon_j = \pi_{ij} \tau_j \quad \text{with } i = 1, 2, 3 \quad \text{and } j = 1, 2, \dots, 6, \quad (47.25)$$

where γ_{ij} and π_{ij} are the elements of the compact matrix of the elastoresistance and piezoresistance coefficients,

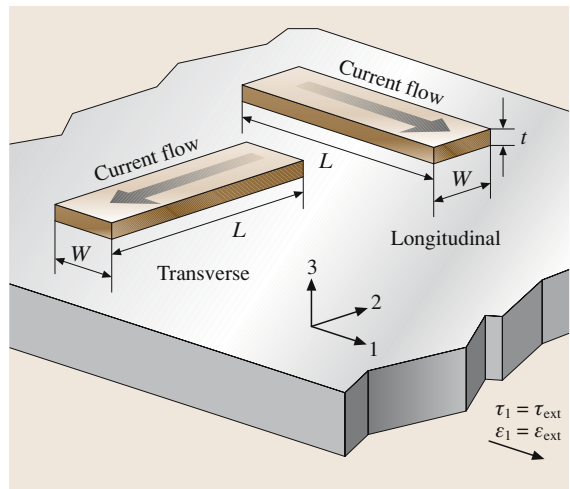


Fig. 47.6 Longitudinal and transverse wires under uniaxial stress and strain components in the 1-direction

respectively, and ϵ_j and τ_j denote the strain and stress components, respectively. (Here, we have used compact notation for these tensors [47.47].)

In metal gauges, the sensitivity of ρ to strain is assumed to be negligible, leading to the well-known longitudinal gauge factor of $k_l = 1 - 2\nu$, where ν is the Poisson's ratio. However, *Arlt* [47.56] has shown that the term $\Delta\rho/\rho$ also includes geometrical attributes. This is due to the change in the volume of the wire and the resulting change in carrier density. According to *Arlt* [47.56], we have for the longitudinal gauge factor

$$k_l = 2 - \frac{\Delta(\gamma\mu)_l}{\epsilon\gamma\mu}, \quad (47.26)$$

where γ denotes the number of free electrons per atom and μ the electron mobility. Since the last term on the right-hand side of (47.26) is relatively small for metallic gauges, the gauge factor mainly stems from geometrical changes; it is close to 2 and temperature independent [47.57].

These results can be validated using beam-deflection (cantilever) experiments performed on strain gauges integrated on glass and silicon substrates. Figure 47.7 illustrates the schematic of a beam-deflection system. One end of the sample is clamped ($x = 0$) and the free end is deflected by a displacement Δ . *Rajanna and Mohan* [47.58] used this method in measurements of both tensile and compressive configurations by simply placing the sample with the films on the top or bottom, respectively (see Fig. 47.7). The strain at location x along a sample with a length L_0 reads

$$\epsilon(x) = \pm \frac{3t_s\Delta}{2L_0^2}(1 - x/L_0), \quad (47.27)$$

with the positive and negative signs denoting tensile and compressive configurations, respectively.

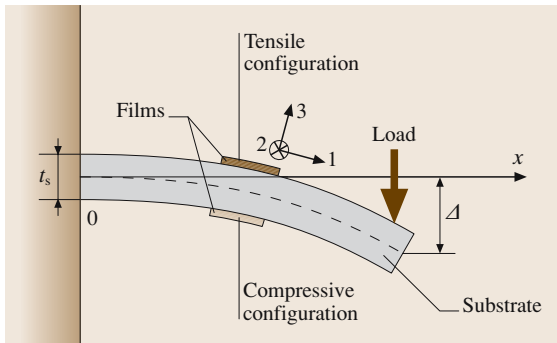


Fig. 47.7 Schematic of a beam-deflection experiment

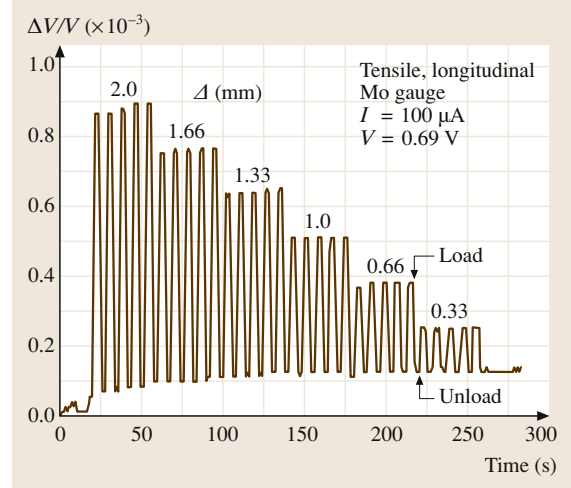


Fig. 47.8 Normalized change in output voltage of a longitudinal Mo strain gauge

Figure 47.8 illustrates the result of the beam-deflection experiment on a 150-nm-thick molybdenum strain gauge with 30 turns, line width of 20 μm , and length of 1100 μm . The strain gauges are biased with a constant current of 100 μA and the voltage drop across its terminals is measured. The gauges are subjected to a sequence of loading/unloading steps in tensile configuration to eliminate systematic errors associated with slowly varying transients. By averaging the value of the voltage modulation, we retrieve a gauge factor of approximately $k_l \approx 2.06$ for the longitudinal strain gauges.

The gauge factor for semiconductors is much higher due to the higher elastoresistance coefficients. *Dössel* [47.57] has related the gauge factor to the elastoresistance coefficients as follows:

$$k_l = \gamma_1 - (\nu_s + \nu_h)\gamma_2 \text{ and } k_t = -\nu_s\gamma_1 + (1 - \nu_h)\gamma_2,$$

where

$$\nu_h = \nu_f \frac{1 - \nu_s}{1 - \nu_f}, \quad (47.28)$$

and $\gamma_1 = 2 - \Delta(\gamma\mu)_l/\gamma\mu\epsilon$ and $\gamma_2 = \Delta(\gamma\mu)_t/\gamma\mu\epsilon$ are the elastoresistance coefficients, and ν_s and ν_f denote the Poisson's ratios for the substrate and the film, respectively.

The same beam-deflection experiment can also be performed on n^+ $\mu\text{c-Si:H}$ strain gauges. Figure 47.9 displays the results of tensile and compressive tests on metal and semiconductor gauges, indicating a higher gauge factor. The values are $k_l = -17.0$ and $k_t = -3.41$

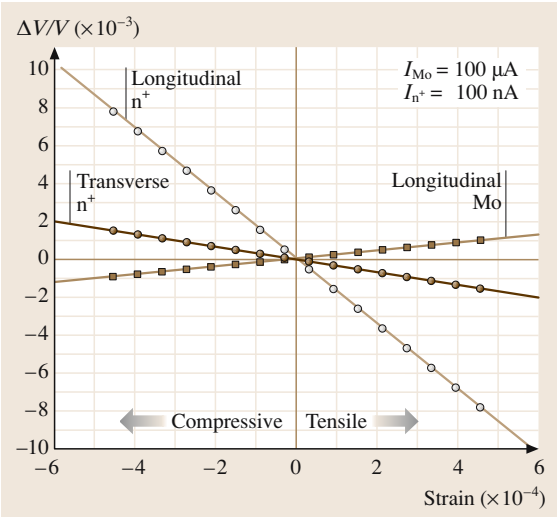


Fig. 47.9 Change in voltage of longitudinal and transverse n^+ $\mu\text{c-Si:H}$ and longitudinal Mo gauges under tensile and compressive strains

for longitudinal and transverse semiconductor gauges, respectively. Assuming $\nu_s = \nu_f = 0.23$, we find that $\gamma_1 = -22$ and $\gamma_2 = -10.9$ for n^+ $\mu\text{c-Si:H}$ films.

The longitudinal gauge factor obtained for the n^+ $\mu\text{c-Si:H}$ gauges corroborate the results of Germer [47.59] for phosphorus-doped microcrystalline thin-film samples with a doping density of $\approx 10^{20} \text{ cm}^{-3}$. For the transverse gauge factor, Germer has observed a small (negative or positive) gauge factor for doping densities in the range $5 \times 10^{19} - 10^{20}$. This shows that the

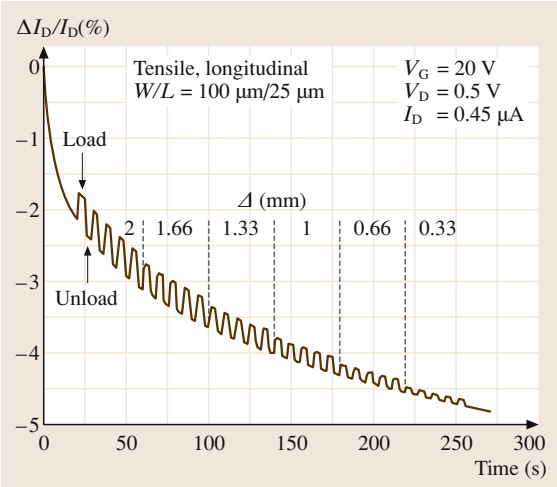


Fig. 47.10 Change in TFT current in deflection experiments

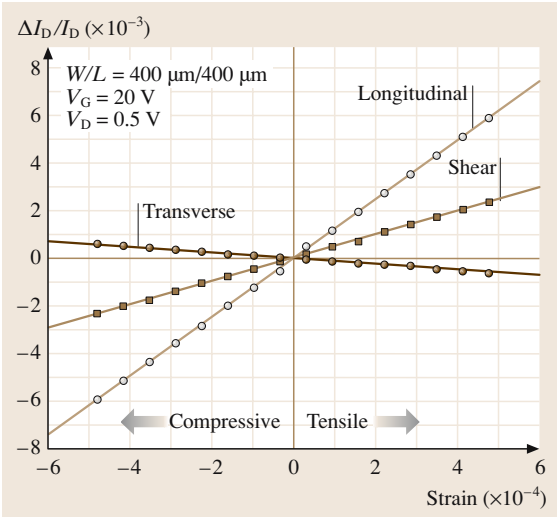


Fig. 47.11 Normalized change in TFT current as a function of tensile and compressive strain

transverse gauge factor is highly sensitive to the doping density and other process conditions, which explains the slight difference between our values and that reported by Germer.

47.3.2 Strained Amorphous-Silicon Transistors

Figure 47.10 illustrates measurement results for the transient drain current of a longitudinal TFT subject to a sequence of tensile loading/unloading steps. We see immediate changes in current superimposed on the intrinsic transient response of the TFT. The measured change in current ΔI_D decreases with decreasing displacement. Correspondingly, we define the sensitivity

$$S_{\epsilon}(I_D) = \frac{1}{I_D} \frac{\partial \Delta I_D}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0}, \quad (47.29)$$

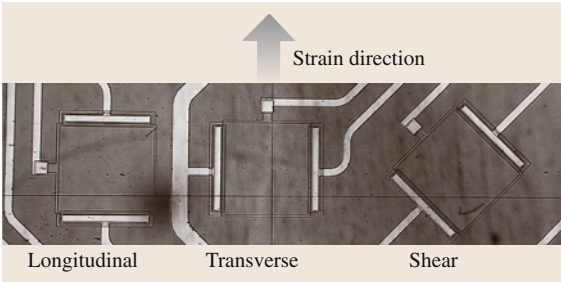


Fig. 47.12 Longitudinal, transverse, and shear TFTs

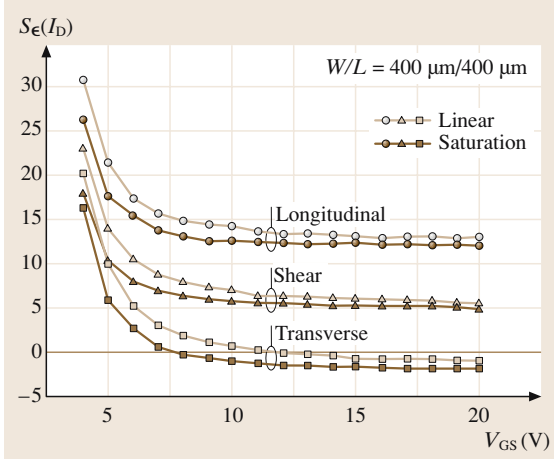


Fig. 47.13 Bias dependence of the TFT current sensitivity

where I_D is the unstrained value of current and ϵ_0 a reference strain value.

Figure 47.11 depicts measurement results for the change in drain current of the longitudinal, shear, and transverse TFTs (Fig. 47.12) under tensile and compressive strains. The TFTs have $W/L = 400 \mu\text{m}/400 \mu\text{m}$ and are biased in the linear regime with constant $V_{GS} = 20 \text{ V}$ and $V_{DS} = 0.5 \text{ V}$. As seen in the figure, the results for compressive and tensile strain are similar but opposite in sign. For the longitudinal TFT, the current increases with tensile strain with a sensitivity $S_\epsilon(I_{D,l}) = 12.5$. In contrast, for the transverse TFT, this is small and negative $S_\epsilon(I_{D,t}) = -1.1$, clearly signifying an orientation dependence. The value of $S_\epsilon(I_{D,s}) = 4.5$ for the shear TFT can be explained from the linear superposition of the effects of longitudinal and transverse strain components.

The measured value for $S_\epsilon(I_{D,l})$ is close to the value of 15 ± 3 reported by Spear and Heintze [47.51] for intrinsic a-Si:H at room temperature. Gleskova et al. [47.60] have found a higher longitudinal sensitivity $S_\epsilon(I_{D,l}) = 26$. For the transverse sensitivity, Spear and Heintze have reported a positive value of $S_\epsilon(I_{D,t}) = 7$ for intrinsic a-Si:H samples.

Table 47.2 Values for $S_\epsilon^H(I_D)$, $S_\epsilon(\mu_{\text{eff}})$, and $S_\epsilon(V_{\text{thr}})$ for different orientations

Parameter	Longi- tudinal	Trans- verse	Shear
$S_\epsilon^H(I_D) = \Delta I_D / I_D \epsilon$	12.1	-1.1	4.5
$S_\epsilon(\mu_{\text{eff}}) = \Delta \mu_{\text{eff}} / \mu_{\text{eff}} \epsilon$	11	-1.1	4.0
$S_\epsilon(V_{\text{thr}}) = \Delta V_{\text{thr}} / V_{\text{thr}} \epsilon$	5	4.5	4.7

In addition to TFT orientation, we observe that the gate bias alters the magnitude of strain-induced change in current ΔI_D . To examine the impact of bias, deflection experiments were performed for different values of the gate bias V_{GS} in the range 4–20 V in 1 V steps. Figure 47.13 illustrates the measured $S_\epsilon(I_D)$ as a function of V_{GS} for TFTs of different orientations. Solid symbols denote measurement data for the linear regime ($V_{DS} = 0.5 \text{ V}$) while the open symbols are those for the saturation regime where the gate and drain terminals are shorted. Interestingly, the modulation in the current shifts toward positive values as the gate bias decreases. This is true for TFTs of all orientations, and independent of whether the devices were integrated on glass or silicon substrates.

However, the sensitivity $S_\epsilon(I_{D,t})$ for the transverse TFT undergoes a sign change. At high biases, the $S_\epsilon(I_{D,t})$ is generally small and negative (i.e. ΔI_D is positive for tensile strain). As the gate bias decreases, the $S_\epsilon(I_{D,t})$ virtually vanishes at approximately 7 V, and subsequently increases to a sizable positive value (i.e. ΔI_D is negative for tensile strain) at lower voltages ($V_{GS} < 7 \text{ V}$).

From the results of Fig. 47.13, we identify two distinct components underlying the strain-induced modulation of current: the high-bias ($V_{GS} > 7 \text{ V}$) $S_\epsilon^H(I_D)$ and low-bias ($V_{GS} < 7 \text{ V}$) $S_\epsilon^L(I_D)$ components such that

$$S_\epsilon(I_D) = S_\epsilon^H(I_D) + S_\epsilon^L(I_D). \quad (47.30)$$

At high biases, $S_\epsilon(I_D)$ of longitudinal, transverse, and shear TFTs gradually approach constant values. As seen from Table 47.2, the extracted values for $S_\epsilon^H(I_D)$ are strongly orientation dependent, suggesting the presence of strain-induced modulation in carrier mobility, whose sensitivity we define as $S_\epsilon(\mu_{\text{eff}}) = \Delta \mu_{\text{eff}} / \mu_{\text{eff}} \epsilon$. The mobility change in the longitudinal orientation is higher than that in the transverse orientation.

Superimposed on the high-bias component is the low-bias component $S_\epsilon^L(I_D)$ which manifests itself as a bias-dependent positive shift in $S_\epsilon(I_D)$. This component can be attributed to the modulation in threshold voltage. Correspondingly, we define the threshold-voltage sensitivity as $S_\epsilon(V_{\text{thr}}) = \Delta V_{\text{thr}} / V_{\text{thr}} \epsilon$. The change in V_{thr} can be attributed to a strain-induced change in the density of deep states, which is orientation independent [47.51]. The modulation in threshold voltage leads to a significant change in current at low biases, and can be quantitatively explained by looking at the current–voltage relation in the linear regime. From (47.16) and using partial differentiation with respect to

strain, we can write $S_\epsilon(I_{DS,lin})$ as

$$S_\epsilon(I_{DS,lin}) = S_G + S_\epsilon(\mu_{eff}) - S_\epsilon(V_{thr}) \frac{\alpha V_{thr}}{V_{GT}} + S_\epsilon(\alpha) \left(\alpha \ln \left| \frac{V_{GT}}{V_0} \right| - \frac{\alpha}{2} - \frac{1}{\alpha-1} \right), \quad (47.31)$$

where $V_0 = (2\epsilon k_B T_i N_0)^{1/2} / C_i$, and

$$S_G = S_\epsilon(W) - S_\epsilon(L) + (\alpha - 1)S_\epsilon(C_i)$$

Here, S_G includes the effect of change in device dimensions. The other terms on the right-hand side of (47.31), in order from the left, describe the dependence of $S_\epsilon(I_{DS,lin})$ on $S_\epsilon(\mu_{eff})$, $S_\epsilon(V_{thr})$, and $S_\epsilon(\alpha)$, respectively, which represent strain-induced modulation of the different TFT parameters (μ_{eff} , V_{thr} , and α).

As seen in (47.31), the modulation in mobility $S_\epsilon(\mu_{eff})$ is directly reflected in the change of current. In contrast, the strain-induced change in threshold voltage $\Delta V_{thr}/V_{thr}$ is scaled, and by a factor of $\alpha V_{thr}/V_{GT}$. This is particularly visible at low biases and its effect decreases with increasing gate bias V_{GT} . Thus, at high biases, the impact of threshold-voltage modulation is minimal, which yields

$$S_\epsilon^H(I_D) = S_G + S_\epsilon(\mu_{eff}). \quad (47.32)$$

In contrast, the low-bias component can be written as

$$S_\epsilon^L(I_D) = -S_\epsilon(V_{thr}) \frac{\alpha V_{thr}}{V_{GT}}. \quad (47.33)$$

It is important to note that the last term on the right-hand side of (47.31) represents the impact of strain-induced

modulation of α . This term contains a bias-dependent scaling factor of $\alpha \ln |V_{GT}/V_0|$, which does not correlate with the observed bias dependence of $S_\epsilon(I_D)$ seen in Fig. 47.13. This implies that modulation in V_{thr} and μ_{eff} are the dominant contributors to the observed changes in current.

The activation energy for the temperature dependence of $S_\epsilon^L(I_D)$ and $S_\epsilon^H(I_D)$ is defined as

$$E_i = \frac{\partial S_\epsilon^i(I_D)}{\partial (1/k_B T)} \text{ with } i = H \text{ or } L. \quad (47.34)$$

The values for E_H and E_L are found to be 140 meV and 0.58 eV, respectively. The much lower activation energy (140 meV) at high biases indicates that the $S_\epsilon^H(I_D)$ stems from the shallow states in the conduction-band tail. Again, this corroborates our previous findings that the sensitivity of the current at high biases is associated with the mobility modulation that is principally determined by the tail states. Spear and Heintze [47.51] have found an activation energy of 0.52 eV for intrinsic and doped a-Si:H layers. This corroborates with our low-voltage sensitivity data. The high activation energy for $S_\epsilon^L(I_D)$ identifies the role of deep states in the gap, which is in agreement with our previous finding that $S_\epsilon(I_D) \propto S_\epsilon(V_{thr})$ at low biases. Here, the strain is believed to modify the energy of the deep states [47.51].

From (47.32) and (47.33), the values of $S_\epsilon(\mu_{eff})$ and $S_\epsilon(V_{thr})$ for different orientations can be determined (Table 47.2). The values can be incorporated in a compact model for the different TFT orientations, which can be used for computer-aided design (CAD) of mechanically flexible TFT circuits [47.61].

References

- 47.1 S. R. Forrest: Nature **428**, 911–918 (2004)
- 47.2 S. Wagner, H. Gleskova, J. C. Sturm, Z. Suo: In: *Technology and Applications of Amorphous Silicon*, ed. by R. A. Street (Springer, Berlin 2000) pp. 222–251
- 47.3 S. E. Shaheen, R. Radspinner, N. Peyghambarian, G. E. Jabbour: Appl. Phys. Lett. **79**, 2996 (2001)
- 47.4 R. A. Street, M. Mulato, R. Lau, J. Ho, J. Graham, Z. Popovic, J. Hor: Appl. Phys. Lett. **78**, 4193–4195 (2001)
- 47.5 P. Servati, Y. Vygranenko, A. Nathan, S. Morrison, A. Madan: J. Appl. Phys. **96**, 7575–7582 (2004)
- 47.6 L. Collins: IEE Rev. **49**(2), 42–45 (2003)
- 47.7 A. B. Chwang, M. A. Rothman, S. Y. Mao, R. H. Hewitt, M. S. Weaver, J. A. Silvernail, K. Rajan, M. Hack, J. J. Brown, X. Chu, L. Moro, T. Krajewsky, N. Rutherford: Appl. Phys. Lett. **83**(3), 413–415 (2003)
- 47.8 A. Nathan, A. Kumar, K. Sakariya, P. Servati, S. Sambandan, D. Striakhilev: IEEE J. Solid-State Circ. **39**(9), 1477–1486 (2004)
- 47.9 M. J. Powell, C. van Berkel, A. R. Franklin, S. C. Deane, W. I. Milne: Phys. Rev. B **45**(8), 4160–4170 (1992)
- 47.10 D. Knipp, R. A. Street, A. Volkel, J. Ho: J. Appl. Phys. **93**, 347 (2003)
- 47.11 S. Alexander, P. Servati, G. R. Chaji, S. Ashtiani, R. Huang, D. Striakhilev, K. Sakariya, A. Kumar, A. Nathan, C. Church, J. Wzorek, P. Arsenaault: J. Soc. Info. Display **13**(7), 587–595 (2005)

- 47.12 G. L. Graff, R. E. Williford, P. E. Burrows: J. Appl. Phys. **96**(4), 1840–1849 (2004)
- 47.13 J. A. Rogers, Z. Bao, A. Dodabalapur, A. Makhija: IEEE Electr. Devices Lett. **21**(3), 100–103 (2000)
- 47.14 C.-S. Yang, L. L. Smith, C. B. Arthur, G. N. Parsons: J. Vac. Sci. Technol. B **18**(2), 683–689 (2000)
- 47.15 H. Gleskova, S. Wagner, V. Gašparík, P. Kováč: Appl. Surf. Sci. **175–176**, 12–16 (2001)
- 47.16 D. Stryahilev, A. Sazonov, A. Nathan: J. Vac. Sci. Technol. A **20**(3), 1087–1090 (2002)
- 47.17 M. Meitine, A. Sazonov: Mat. Res. Soc. Symp. Proc. **769**, H6.6.1 (2003)
- 47.18 W. A. MacDonald: J. Mater. Chem. **14**, 4–10 (2004)
- 47.19 K. L. Chopra: *Thin Film Phenomena* (McGraw-Hill, Toronto 1969)
- 47.20 E. Y. Ma, S. Wagner: Appl. Phys. Lett. **74**, 2661–2662 (1999)
- 47.21 A. Madan, P. G. Le Comber, W. E. Spear: J. Non-Cryst. Solids **20**, 239–257 (1976)
- 47.22 N. F. Mott, E. A. Davis: *Electronic Processes in Non-Crystalline Materials* (Oxford University Press, Oxford 1971)
- 47.23 M. H. Cohen, H. Fritzsche, S. R. Ovshinsky: Phys. Rev. Lett. **22**(20), 1066–1068 (1969)
- 47.24 C. Popescu, T. Stoica: In: *Thin Film Resistive Sensors*, ed. by P. Ciureanu, S. Middelhoek (IOPP, New York 1992) pp. 37–112
- 47.25 P. W. Anderson: Phys. Rev. **109**(5), 1492–1505 (1958)
- 47.26 D. Monroe: Phys. Rev. Lett. **54**(2), 146–149 (1985)
- 47.27 R. A. Street: *Hydrogenated Amorphous Silicon* (Cambridge University Press, New York 1992)
- 47.28 W. E. Spear, P. G. Le Comber: J. Non-Cryst. Solids **8–10**, 727–738 (1972)
- 47.29 M. J. Powell: Philos. Mag. B **43**(1), 93–103 (1981)
- 47.30 C.-Y. Huang, S. Guha, S. J. Hudgens: Phys. Rev. B **27**(12), 7460–7465 (1983)
- 47.31 J. D. Cohen, D. V. Lang, J. P. Harbison: Phys. Rev. Lett. **45**(3), 197–200 (1980)
- 47.32 M. Hirose, T. Suzuki, G. H. Döhler: Appl. Phys. Lett. **34**(3), 234–236 (1979)
- 47.33 P. Viktorovitch, G. Moddel: J. Appl. Phys. **51**(9), 4847–4854 (1980)
- 47.34 M. Shur, M. Hack: J. Appl. Phys. **55**(10), 3831–3842 (1984)
- 47.35 A. Salleo, T. W. Chen, A. R. Völkel, Y. Wu, P. Liu, B. S. Ong, R. A. Street: Phys. Rev. B **70**, 115–311 (2004)
- 47.36 M. J. C. M. Vissenberg, M. Matters: Phys. Rev. B **57**(20), 12964–12967 (1998)
- 47.37 J. G. Shaw, M. Hack: J. Appl. Phys. **64**(9), 4562–4566 (1988)
- 47.38 C. Tanase, E. J. Meijer, P. W. M. Blom, D. M. de Leeuw: Phys. Rev. Lett. **91**(21), 216–601 (2003)
- 47.39 T. Tiedje, A. Rose: Solid State Comm. **37**, 49–52 (1980)
- 47.40 B. S. Ong, Y. Wu, P. Liu, S. Gardner: J. Am. Chem. Soc. **126**, 3378–3379 (2004)
- 47.41 A. J. Campbell, M. S. Weaver, D. G. Lidzey, D. D. C. Bradley: J. Appl. Phys. **84**(12), 6737–6746 (1998)
- 47.42 D. Natali, M. Sampietro: J. Appl. Phys. **92**(9), 5310–5318 (2002)
- 47.43 P. Servati, D. Striakhilev, A. Nathan: IEEE Trans. Electr. Devices **50**(11), 2227–2235 (2003)
- 47.44 C. Tanase, E. J. Meijer, P. W. M. Blom, D. M. de Leeuw: Organic Elect. **4**, 33 (2003)
- 47.45 F. R. Shapiro, D. Adler: J. Non-Cryst. Solids **74**(2–3), 189–194 (1985)
- 47.46 S. D. Baranovskii, T. Faber, F. Hensel, P. Thomas: J. Phys.: Condens. Matter **9**, 2699 (1997)
- 47.47 A. Nathan, H. Baltes: *Microtransducer CAD* (Springer, Wien 1999)
- 47.48 C. Herring, E. Vogt: Phys. Rev. **101**(3), 944–961 (1956)
- 47.49 V. A. Gridchin, V. M. Lubimsky, M. P. Sarina: Sens. Actuators A **49**(1–2), 67–72 (1995)
- 47.50 A. Dévényi, A. Belu, G. Korony: J. Non-Cryst. Solids **4**, 380–390 (1970)
- 47.51 W. E. Spear, M. Heintze: Philos. Mag. B **54**(5), 343–358 (1986)
- 47.52 B. Welber, M. H. Brodsky: Phys. Rev. B **16**(8), 3660–3664 (1977)
- 47.53 B. A. Weinstein: Phys. Rev. B **23**(2), 787–793 (1981)
- 47.54 R. Zallen, W. Paul: Phys. Rev. **155**(3), 703–711 (1967)
- 47.55 D. Lazarus: Phys. Rev. B **24**(4), 2282–2284 (1981)
- 47.56 G. Arlt: J. Appl. Phys. **49**(7), 4273–4274 (1978)
- 47.57 O. Dössel: Sens. Actuators **6**(3), 169–179 (1984)
- 47.58 K. Rajanna, S. Mohan: Phys. Status Solidi A **105**(2), K181–K184 (1988)
- 47.59 W. Germer: Sens. Actuators **7**(2), 135–142 (1985)
- 47.60 H. Gleskova, S. Wagner, W. Soboyejo, Z. Suo: J. Appl. Phys. **92**(10), 6224–6229 (2002)
- 47.61 P. Servati, A. Nathan: Appl. Phys. Lett. **86**(7), 033504 (2005)

48. Photoconductors for X-Ray Image Detectors

Recent flat-panel X-ray imaging detectors have been shown to be able to replace present-day X-ray film/screen cassettes; they capture an X-ray image electronically and hence enable a clinical transition to digital radiography. This chapter critically discusses the material, transport and imaging detector properties (e.g., dark current) of several potential X-ray photoconductors and compares them with an ideal photoconductor for use in direct-conversion imaging detectors. The present chapter also considers various metrics of detector performances including sensitivity, detective quantum efficiency, resolution in terms of the modulation transfer function, image lag and ghosting; and examines how these metrics

48.1 X-Ray Photoconductors	1123
48.1.1 Ideal Photoconductor Properties	1123
48.1.2 Potential Photoconductors	1124
48.1.3 Summary and the Future	1130
48.2 Metrics of Detector Performance	1131
48.2.1 X-Ray Sensitivity	1131
48.2.2 Detective Quantum Efficiency.....	1133
48.2.3 Modulation Transfer Function (MTF)	1134
48.3 Conclusion	1136
References	1136

depend on the photoconductor, and detector structure and design.

The flat-panel X-ray image detector (sensor) concept is illustrated in Fig. 48.1 where X-rays passing through the patient's hand are incident on a sensor. The sensor converts the incident image to a digital image that can be sent to a computer and viewed on a monitor. Research over the past ten years [48.1] has indicated that the most practical flat-panel medical digital radiographic systems (mammography, chest radiography and fluoroscopy) are those based on a large-area inte-

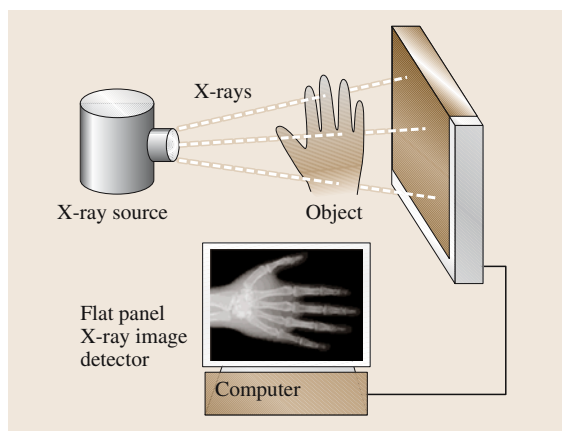


Fig. 48.1 Schematic illustration of an active matrix flat-panel imaging (AMFPI) system used for X-ray imaging

grated circuit or active matrix array (AMA). Flat-panel imagers incorporating active matrix arrays are called active matrix flat-panel imagers or AMFPI. Active matrix arrays using hydrogenated amorphous silicon (a-Si:H) thin-film transistors (TFTs) have been shown to be practical pixel-addressing systems for displays [48.2]. These a-Si:H arrays can be converted into X-ray-sensitive imaging devices by adding a thick (0.1–1 mm) X-ray-detecting medium, either a phosphor or a photoconductor. The physical form of the X-ray AMFPI is similar to a film/screen cassette and thus it will easily fit into current medical film/screen-based X-ray imaging systems. The X-ray image is stored and displayed on the computer almost immediately after the X-ray exposure. The stored image can be rapidly transmitted to remote locations for consultation and analysis. The dynamic range of recently developed AMFPI systems is much higher than film/screen-based imaging systems [48.3]. AMFPI are able to read out an entire image in 1/30 s, sufficiently rapid to perform fluoroscopy (real-time imaging) [48.4].

The active matrix array used for pixel addressing and readout in an AMFPI consists of many single pixels, each of which represents a corresponding pixel of the image. Each pixel collects charge proportional to the X-ray radiation that it receives. To generate this signal charge, either a phosphor converts the X-rays

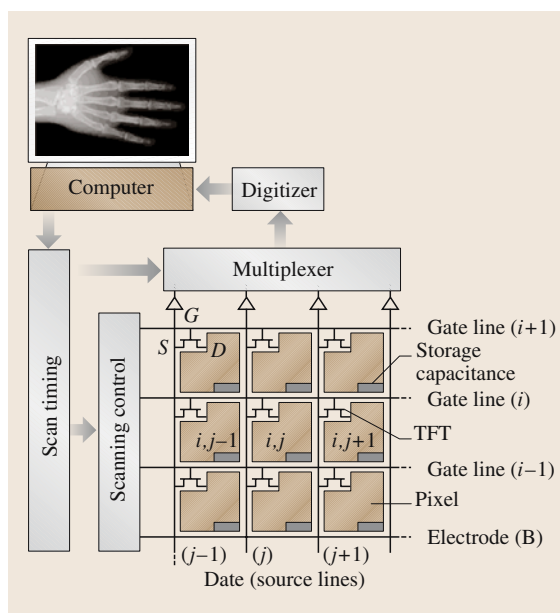


Fig. 48.2 Schematic diagram that shows some pixels of the active matrix array for use in X-ray AMFPI with self-scanned electronic readout. The charge distribution residing on the panel's pixels are read out by scanning the arrays row by row using peripheral electronics and multiplexing the parallel columns into a serial digital signal

to visible light, which in turn is detected with a pin photodiode at the pixel (indirect) or an X-ray photoconductor converts the incident X-rays to charge (direct) in one step. For both the indirect and direct conversion approaches, the latent image is a charge distribution residing on the array's pixels. The charges are read out by scanning the arrays row by row using peripheral electronics and multiplexing the parallel columns into a serial digital signal. This signal is then transmitted to a computer system for storage and display. Several manufacturers and academic researchers have used the indirect approach [48.5, 6]. However, we believe that the direct approach, due to its higher conversion gain, has the potential to produce systems superior in image quality to indirect conversion sensors and be both easier and cheaper to manufacture due to their simpler structure. This chapter considers only the direct-conversion X-ray imager and how its dark current, sensitivity, resolution and detective quantum efficiency (DQE) depend on the photoconductor and detector structure. This chapter also discusses essential photoconductor properties, charge transport and imaging properties of promising photoconductors.

Active matrix arrays allow a monolithic imaging system of large area (e.g. 40 cm × 40 cm) to be constructed. As for conventional integrated circuits, planar processing of the array through deposition and doping of lithographically masked individual layers of metals, insulators and semiconductors implement the design of active matrix arrays. In the future, even larger areas should become feasible if required. Millions of individual pixel electrodes in the matrix and are connected, as shown in Fig. 48.2. Each pixel has its own thin-film transistor (TFT) switch and storage capacitor to store image charges. The TFT switches control the image charge so that one line of pixels is activated electronically at a time. Normally, all the TFTs are turned off, permitting the latent image charge to accumulate on the array. The readout is achieved by external electronics and software controlling the state of the TFT switches. The active matrix array consists of $M \times N$ (e.g. 2480 × 3072) storage capacitors C_{ij} whose stored image charge can be read through properly addressing the TFT (i, j) via the gate (i) and source (j) lines. The charges read on each C_{ij} are converted to a digital image. The image readout technique is described by Rowlands and Kasap [48.7]. The readout is self-scanning in that no external means such as a laser are used. The scanning is part of the AMFPI electronics and software, thus permitting a truly compact device.

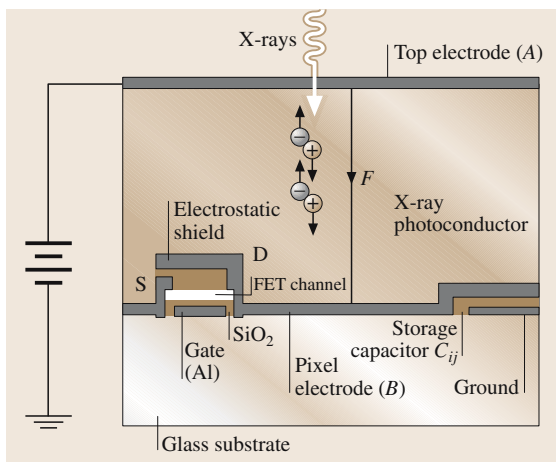


Fig. 48.3 Cross section of a single pixel (i, j) with a TFT switch. The top electrode (A) on the photoconductor is a vacuum-coated metal (e.g. Al). The bottom electrode (B) is the pixel electrode that is one of the plates of the storage capacitance (C_{ij}). (Not to scale—the TFT height is highly exaggerated)

To construct a direct-conversion X-ray AMFPI, a large-band-gap (> 2 eV) high-atomic-number semiconductor or X-ray photoconductor (e.g., stabilized amorphous selenium, a-Se) layer is coated onto the active matrix array. An electrode (A) is subsequently deposited onto the photoconductor layer to facilitate the application of a biasing potential and, hence, an electric field F in the photoconductor layer as shown in Fig. 48.3. The biasing potential applied to the radiation receiving electrode A (top electrode) may be positive or negative, the selection of which depends on many factors discussed later in this chapter. The applied bias varies from a few hundred to several thousand Volts. The capacitance C_{pc} of the photoconductor layer over the pixel is in series with, and much smaller than, the pixel capacitance C_{ij} attached to the pixel electrode, so that most of the applied potential is dropped across the photoconductor and not across the pixel capacitance. The electron-hole pairs (EHPs) generated in the photoconductor by the absorption of X-ray photons travel along the field lines and are collected by the electrodes. If the applied bias voltage is positive, then electrons collect at the positive bias electrode and holes accumulate on the storage capacitor C_{ij} . Each pixel electrode carries an amount of charge Q_{ij} proportional to the

amount of incident X-ray radiation in the photoconductor layer over that pixel, which can be read during self-scanning.

The selection of the photoconductor material for use in direct-conversion X-ray image detectors is currently an important research field in electronic materials. There are various competing semiconductors, such as amorphous (a-)Se, HgI_2 , $CdZnTe$, PbI_2 , PbO and $TlBr$. Detectors based on a-Se have already been developed and some have been commercialized for mammography, general radiography and fluoroscopy [48.8] with active areas as large as $43\text{ cm} \times 43\text{ cm}$. This chapter critically discusses charge-transport and imaging-detector properties (e.g., dark current and image lag) of these photoconductors and compares them with an ideal photoconductor for X-ray imaging detectors. The present chapter also examines various imaging characteristics of photoconductor-based X-ray AMFPIs, including sensitivity (S), detective quantum efficiency (DQE), resolution in terms of the modulation transfer function (MTF), image lag and ghosting. We examine how these characteristics depend not only on the photoconductor's charge-transport properties but also on the detector structure, i.e., the size of the pixel and the thickness of the photoconductor.

48.1 X-Ray Photoconductors

48.1.1 Ideal Photoconductor Properties

The performance of direct-conversion X-ray detectors depends critically on the selection and design of the photoconductor. It is therefore instructive to identify what constitutes a nearly ideal X-ray photoconductor to guide a search for improved performance or a better material. Ideally, the photoconductive layer should possess the following material properties:

1. Most of the incident X-ray radiation should be absorbed within a practical photoconductor thickness to avoid unnecessary patient exposure. This means that, over the energy range of interest, the absorption depth δ of the X-rays must be substantially less than the device layer thickness L . In other words, the quantum efficiency (η) should be high.
2. The photoconductor should have high intrinsic X-ray sensitivity, i.e., it must be able to generate as many collectable (free) electron-hole pairs (EHPs) as possible per unit of incident radiation. This means

the amount of radiation energy required, denoted by W_{\pm} , to create a single free electron-hole pair must be as low as possible. Typically, W_{\pm} increases with the band gap E_g of the photoconductor [48.9] and thus a low E_g is desired for maximum X-ray sensitivity.

3. There should be little bulk recombination of electrons and holes as they drift to the collection electrodes; EHPs are generated in the bulk of the photoconductor. Bulk recombination is proportional to the product of the concentration of holes and electrons, and typically it is negligible for clinical exposure rates (i.e. provided the instantaneous X-ray exposure is not too high).
4. There should be negligible deep trapping of EHPs, which means that, for both electrons and holes, the schubweg defined as $\mu\tau F \gg L$ where μ is the drift mobility, τ is the deep-trapping time (lifetime), F is the electric field and L is the photoconductor layer thickness. The schubweg is the mean distance a carrier drifts before it is trapped and unavailable for

- conduction. The temporal responses of the X-ray image detector, such as lag and ghosting, depend on the rate of carrier trapping.
5. The diffusion of carriers should be negligible compared with their drift. This property ensures less time for lateral carrier diffusion and leads to better spatial resolution.
 6. The dark current should be as small as possible, because it is a source of noise. This means the contacts to the photoconductor should be non-injecting and the rate of thermal generation of carriers from various defects or states in the band gap should be negligibly small (i. e. dark conductivity is practically zero). Small dark conductivity generally requires a wide-band-gap semiconductor, which conflicts with condition (2) above. The dark current should preferably not exceed 10–1000 pA/cm², depending on the clinical application [48.10].
 7. The longest carrier-transit time, which depends on the smallest drift mobility, must be shorter than the image readout time and inter-frame time in fluoroscopy.
 8. The properties of the photoconductor should not change with time because of repeated exposure to X-rays, i. e. X-ray fatigue and X-ray damage should be negligible.
 9. The photoconductor should be easily coated onto the AMA panel (typically 30×30 cm and larger), for example, by conventional vacuum techniques without raising the temperature of the AMA to damaging levels (e.g., ≈ 300 °C for a-Si panels). This eliminates the possibility of using single-crystal materials that would require extended exposure to much higher temperature if they were to be grown directly onto the panel.
 10. The photoconductor should have uniform characteristics over its entire area.
 11. The temporal artifacts such as image lag and ghosting should be as small as possible. Lag is the carryover of image charge generated by previous X-ray exposures into subsequent image frames. The residual signal fractions following a pulsed X-ray irradiation are referred to as “image lag”. Ghosting is the change of X-ray sensitivity of the X-ray image detector as a result of previous exposure to radiation. In the presence of ghosting, a shadow impression of a previously acquired image is visible in subsequent uniform exposures.

The large-area-coating requirement in (9) rules out the use of crystalline semiconductors, whose only practi-

cal production process is to grow large boules, which are subsequently sliced. Thus, only amorphous or polycrystalline (poly) photoconductors are currently practical for use in large-area X-ray imaging detectors. Amorphous selenium (a-Se) is one of the most highly developed photoconductors for large-area detectors due to its commercial use in photocopiers and laser printers as an electrophotographic photoreceptor [48.11]. In fact, the direct-conversion flat-panel imaging technology has been made possible by the use of two key elemental amorphous semiconductors: a-Si:H (used for TFTs) and a-Se (used for photoconductor layers). Although their properties are different, both can be readily prepared in large areas, which is essential for an X-ray imaging detector. Stabilized a-Se (a-Se alloyed with 0.2–0.5% As and doped with 10–40 ppm Cl) is currently the preferred photoconductor for clinical X-ray image sensors, because it can be quickly and easily deposited as a uniform film over large areas and it has an acceptable X-ray absorption coefficient, good charge-transport properties for both holes and electrons, and lower dark current than many competing polycrystalline layers [48.9, 10]. Flat-panel X-ray image detectors with an a-Se photoconductor have been shown to provide excellent images.

There has been active research to find potential X-ray photoconductors to replace a-Se in flat-panel image detectors because of the substantially higher W_{\pm} and operating electric field of a-Se compared to other potential X-ray photoconductors [48.10, 12]. For example, the typical value of the electric field used in a-Se devices is 10 V/μm where the value of W_{\pm} is about 45 eV; the value of W_{\pm} is 5–6 eV for polycrystalline mercuric iodide (poly-HgI₂) and polycrystalline CdZnTe. The main drawback of polycrystalline materials is the adverse effects of grain boundaries in limiting charge transport and the nonuniform response of the sensor due to large grain sizes. Grain boundaries in the polycrystalline material are expected to create trapping levels within the band gap and introduce potential barriers between neighboring grains [48.13]. Another disadvantage of these polycrystalline detectors is the higher dark current compared to a-Se detectors. However, there have been efforts to improve the material properties and reduce the dark currents of polycrystalline HgI₂ and poly-CdZnTe-based image detectors [48.14, 15]. Recent experiments on large-area HgI₂, PbI₂, CdZnTe (< 10% Zn), and PbO polycrystalline X-ray photoconductive layers deposited on active matrix arrays have shown encouraging results [48.16–18]. A more detailed description of these three potential photoconductors for

direct-conversion AMFPIs is presented below along with TlBr.

48.1.2 Potential Photoconductors

The properties of an ideal photoconductor for X-ray image detectors have been discussed in the section above. In this section, important properties of potential photoconductors for X-ray image detectors are discussed and compared with the ideal case.

Amorphous Selenium (a-Se)

Stabilized a-Se can be easily coated as thick films (e.g., 100–1000 μm) onto suitable substrates by conventional vacuum deposition techniques and without the need to raise the substrate temperature beyond 60–70°C (much below the damage threshold of the AMA, e.g., $\approx 300^\circ\text{C}$ for a-Si:H panels). Its amorphous state maintains uniform characteristics to a very fine scale over large areas. Stabilized a-Se, not pure a-Se, is used in X-ray sensors, because pure a-Se is thermally unstable and crystallizes over several weeks or months following manufacture. Crystalline Se is unsuitable as an X-ray photoconductor because it has a much lower dark resistivity and hence orders of magnitude larger dark current

than a-Se. Alloying pure a-Se with As (0.2–0.5% As) greatly improves the stability of the composite film and helps to prevent crystallization. However, it is found that the addition of As has an adverse effect on the hole lifetime because the arsenic introduces deep hole traps. If the alloy is doped with 10–20 parts per million (ppm) of a halogen (such as Cl), the hole lifetime is restored to its initial value. Thus, a-Se film that has been alloyed with 0.2–0.5% As (nominal 0.3% As) and doped with 10–20 ppm Cl is called stabilized a-Se. The density of a-Se is 4.3 g/cm^3 with an energy gap of $E_g \approx 2.1\text{--}2.2\text{ eV}$.

There are localized states in the so-called forbidden gap of a-Se. Some of these are located near the band edges (shallow traps), while some are located deep in the energy band (deep traps). Localized states are simply traps and are not extended throughout the material, but are localized in space (both shallow and deep) due to various structural defects that are stable at room temperature. Drift of both electrons and holes involves interactions with shallow and deep traps, as shown in Fig. 48.4. Shallow traps reduce the drift mobility; deep traps prevent the carriers from crossing the photoconductor. The effective drift mobility μ of carriers is the mobility μ_0 in the extended states reduced by the trapping and release events due to the presence of shallow traps,

$$\mu = \frac{\tau_c}{\tau_c + \tau_r} \mu_0, \quad (48.1)$$

where τ_c and τ_r are the average capture and release times in the shallow-trap centers. The capture time represents the mean time that a mobile carrier drifts in the extended states before becoming trapped in a shallow-trap center. The release time is the mean time that a carrier remains in a trap before being released back into the extended states. Re-emission from a shallow trap is mostly dominated by thermally activated processes. The shallow-trap release time is relatively short and a typical carrier may experience many shallow capture and release events while traversing the detector thickness (100–1000 μm).

Although the nature of the shallow traps in a-Se has not been fully established, the drift mobilities of both holes and electrons are quite reproducible. The room-temperature effective hole mobility μ_h is independent of the preparation of the sample and has a value of $\approx 0.12\text{ cm}^2/\text{Vs}$, whereas the effective electron mobility μ_e is in the range $0.003\text{--}0.006\text{ cm}^2/\text{Vs}$ [48.12]. The hole drift mobility does not change with the addition of As or Cl. The value of μ_e decreases with the addition of

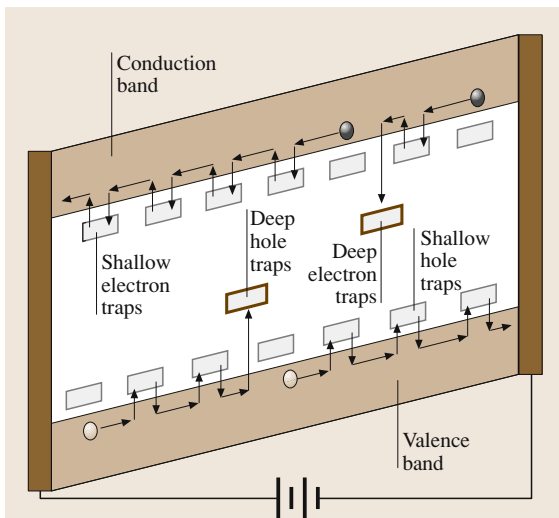


Fig. 48.4 Diagram illustrating the band gap of a photoconductor with an applied electric field that tilts the bands, encouraging drift of holes in the direction of the field and electrons counter to the field. Drift of both electrons and holes involves interactions with shallow and deep traps. Shallow traps reduce the drift mobility and deep traps prevent the carriers from crossing the photoconductor

As to a-Se (e.g., in stabilized a-Se) but Cl doping does not affect it.

Once a carrier is caught in a deep trap, it will remain immobile until a lattice vibration gives it enough energy to be excited back into the extended states, where it can drift once again. The deep-trap release time is very long (minutes to hours), and a deeply trapped carrier is essentially permanently removed from conduction. Therefore, the carrier lifetime depends on the concentration of deep rather than shallow traps. The charge-carrier lifetimes vary substantially between different samples and depend on factors such as the source of a-Se material, impurities, and the preparation method. The electron lifetime τ_e is particularly sensitive to impurities in the a-Se source material. The hole lifetime τ_h drops rapidly with decreasing substrate temperature (temperature of the a-Se substrate during the evaporation process) whereas τ_e is unaffected. Increasing the As concentration in a-Se decreases τ_h and increases τ_e [48.19]. On the other hand, Cl doping increases τ_h and decreases τ_e . The typical lifetimes in stabilized a-Se are in the range 10–500 μ s for holes and 100–1500 μ s for electrons [48.12].

The fractional increase in the τ_e with As addition is greater than the drop in μ_e . Thus the electron range ($\mu_e \tau_e$) increases with As content. The effect of Cl doping on the carrier ranges ($\mu \tau$ products) is more pronounced than that of As doping. Thus, we can control both electron and hole ranges by appropriately choosing the relative amounts of As and Cl in a-Se.

The electron-hole-pair creation energy W_{\pm} in a-Se has a strong dependence on electric field F but only a weak dependence on the X-ray photon energy E [48.20, 21]. The quantity W_{\pm} is decreased by increasing either F or E . W_{\pm} at a given E in a-Se follows an empirical relation given by

$$W_{\pm} \approx W_{\pm}^0 + \frac{B(E)}{F^n}, \quad (48.2)$$

where $B(E)$ is a constant that depends on E , W_{\pm}^0 is the saturated EHP creation energy (at infinite F), and n is typically in the range 0.7–1 [48.22]. The value of W_{\pm}^0 should be $2.2E_g + E_{\text{phonon}}$ [48.23], where E_{phonon} is the phonon energy. With $E_g \approx 2.1$ eV and $E_{\text{phonon}} < 0.5$ eV, we would expect that $W_{\pm}^0 \approx 5$ eV. The energetic primary electron generates many EHPs but only a certain fraction of these are free to drift and the rest recombine before they can contribute to the photocurrent. There are various possible explanations for the F dependence of the EHP creation energy. First, the simultaneously generated electron and its hole twin are attracted to each other by their mutual Coulombic force and may even-

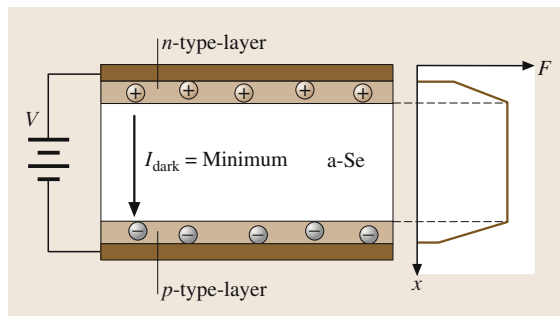


Fig. 48.5 A multilayer pin-type a-Se device structure for blocking dark current

tually recombine. This type of recombination is called geminate recombination (from Gemini—the twins). Another possible mechanism is columnar recombination that involves the recombination of nongeminate electrons and holes generated close to each other in the columnar track of the single high-energy electron (primary) created by the absorption of an X-ray. In both the geminate and columnar cases, the number of carriers escaping recombination should increase with increasing F , which acts to separate the oppositely charged carriers. The question of whether the F dependence of W_{\pm} in a-Se is dominated by geminate or columnar recombination has not been fully resolved and is currently an area of research [48.23, 24]. However, the energy dependence of W_{\pm} is better understood. It decreases slowly with increasing photon energy in the diagnostic [48.20] and megavolt range [48.25]. The total change in W_{\pm} from 20 keV to 6 MeV is of the order of three. This appears to be due to a reduction of recombination with increase in energy. The rate of deposition of energy per unit distance traveled by a primary electron decreases as a function of energy, decreasing the density of EHPs in the column around it. This is expected to reduce columnar recombination—as is seen. Thus it appears that at low energies the contribution from columnar recombination is approximately twice that from geminate, but at high enough energy the columnar effect is reduced to zero. The typical value of the F used in a-Se devices is 10 V/ μ m where the value W_{\pm} is 35–55 eV over the diagnostic beam energy (12–120 keV) and ≈ 15 eV at megavolt energies.

The dark resistivity of a-Se is $\approx 10^{14}$ Ω cm. The bulk-induced dark current in a-Se detectors is within the acceptable limit (< 1 nA/cm²) for $F \leq 20$ V/ μ m. However, the injected dark current for metal/a-Se/metal single-layer structures depends on the nature of the metal/a-Se contacts and nonlinearly on F . It has

however been possible to reduce the dark current to a negligible level in multilayer a-Se structures, e.g., ($< 0.1 \text{ nA/cm}^2$ at F as high as $20 \text{ V}/\mu\text{m}$ [48.26]) by use of blocking contacts. A pin (p-type–intrinsic–n-type) a-Se structure is shown in Fig. 48.5, where the thickness of both the p and n blocking contact layers is a few microns. These p- and n-layers are appropriately doped to serve as unipolar conducting layers that easily trap electrons and holes, respectively, but allow the transport of oppositely charged carriers. The rate of emission of these deeply trapped carriers is so small that there is no significant current injection into the bulk a-Se layer. With these n- and p-layers, the F at the metal electrodes is sufficiently small to minimize charge injection from the contacts, which substantially reduces the dark current.

The image lag in a-Se detectors is under 2% after 33 ms and less than 1% after 0.5 s in the fluoroscopic mode of operation [48.27]. Therefore, image lag in a-Se detectors is considered to be negligible. The pixel-to-pixel sensitivity variation is also negligible in a-Se detectors. The presampling MTF of these detectors is almost equal to the theoretical MTF (sinc function) determined by the pixel aperture [48.28].

Mercuric Iodide (HgI_2)

Poly- HgI_2 has been used as a photoconductor layer in X-ray image detectors. It has been prepared by either physical vapor deposition (PVD) or screen printing (SP) from a slurry of HgI_2 crystals using a wet particle-in-binder process [48.16]. There appears to be no technological barrier to preparing large-area layers, and direct-conversion X-ray AMFPI of $20 \times 25 \text{ cm}^2$ (1536×1920 pixels) and $5 \times 5 \text{ cm}^2$ (512×512 pixels) size have been demonstrated using PVD [48.16] and SP poly- HgI_2 layers. The prototype HgI_2 sensors can potentially be used for mammographic, fluoroscopic or radiographic imaging. There has been active research to improve the material properties of poly- HgI_2 -based image detectors including improving the nonuniformity by reducing the grain size. The band-gap energy $E_g = 2.1 \text{ eV}$, the ionization energy $W_{\pm} \approx 5 \text{ eV}$ and the density of poly- HgI_2 is 6.3 g/cm^3 . The resistivity of this material is $\approx 4 \times 10^{13} \Omega \text{ cm}$ [48.29].

The HgI_2 layer is deposited onto either conductive [indium-tin-oxide- (ITO) or gold-coated] glass plates or a-Si TFT arrays. HgI_2 tends to react chemically with most metals; hence a thin blocking layer (typically, $\approx 1 \mu\text{m}$ layer of insulating polymer) is used between the HgI_2 layer and the pixel electrodes to prevent the reaction and also to reduce the dark current. The HgI_2 layer thickness varies in the range

$100\text{--}500 \mu\text{m}$ and the grain size is $20\text{--}60 \mu\text{m}$. Several hundred angstroms of palladium (Pd) or Au are deposited (by direct evaporation) on top of the HgI_2 layer followed by a polymer encapsulation layer to form a bias electrode.

The dark current of HgI_2 imagers increases super-linearly with the applied bias voltage. For PVD HgI_2 detectors the dark current depends strongly on the operating temperature. It increases by a factor of about two for each 6°C of temperature rise. It is reported that the dark current varies from $\approx 2 \text{ pA/mm}^2$ at 10°C to $\approx 180 \text{ pA/mm}^2$ at 35°C at $F = 0.95 \text{ V}/\mu\text{m}$, which is not desirable for medical imagers (the maximum dark current for medical imaging should be $< 10 \text{ pA/mm}^2$). Therefore, PVD HgI_2 imagers should be operated at relatively low bias (preferably less than $\approx 0.5 \text{ V}/\mu\text{m}$) and relatively low temperature ($< 25^\circ\text{C}$). The dark current in the SP sample is an order of magnitude smaller than in the PVD sample and more stable against temperature variation. It is possible to keep the dark current below 10 pA/mm^2 at temperatures up to 35°C and $F = 1 \text{ V}/\mu\text{m}$. A major disadvantage of SP detectors is that they show 2–4 times less sensitivity compared to PVD detectors.

Electrons have much longer range than holes in HgI_2 and thus the receiving electrode is negatively biased to obtain a higher sensitivity. The $\mu_e\tau_e$ for electrons in the SP HgI_2 is in the range $10^{-6}\text{--}10^{-5} \text{ cm}^2/\text{V}$ [48.16, 30], and $\mu_e\tau_e$ in the PVD sample is about an order of magnitude greater. Recently it was reported that $\mu_e\tau_e$ in PVD HgI_2 is in the range $10^{-5}\text{--}10^{-4} \text{ cm}^2/\text{V}$, which is almost equal to that of single-crystal HgI_2 [48.14, 31]. The reason is that the PVD HgI_2 layer grows in a columnar structure perpendicular to the substrate. Thus a charge carrier may drift along a column without having to pass through grain boundaries, where it would encounter excess trapping and/or recombination. Samples with larger grain sizes may have fewer grain-boundary defects. Hence there is a trend of increasing $\mu_e\tau_e$ with grain size in the PVD samples. However, this trend is not observed in the SP sample, which indicates another mechanism is responsible for its low $\mu_e\tau_e$. Larger grain sizes may cause nonuniform sensor response. The grain sizes must be much smaller than the pixel size to get a uniform response. $\mu_h\tau_h$ in poly- HgI_2 is of the order $\approx 10^{-7} \text{ cm}^2/\text{V}$ [48.30].

Two important drawbacks of polycrystalline sensors are the image lag and the pixel-to-pixel sensitivity variation or nonuniform response. The lowest image-lag characteristics reported are $\approx 7\%$ first-frame lag, $\approx 0.8\%$ after 1 s and $\approx 0.1\%$ at 3 s in fluoroscopic

mode (15 frames/s). The pixel-to-pixel sensitivity variation reduces the dynamic range of the imagers. The relative standard deviation of the sensitivity (standard deviation/average value) in the latest HgI₂ AMFPI is $\approx 10\%$. It is reported that HgI₂ image detectors with smaller grain sizes show good sensitivity and also an acceptably uniform response. The presampling MTF of these detectors is almost equal to the theoretical MTF (sinc function) determined by the pixel aperture.

Cadmium Zinc Telluride (CdZnTe)

CdZnTe ($< 10\%$ Zn) polycrystalline film has been used as a photoconductor layer in X-ray AMFPI. CdZnTe is commonly called CZT. Although CZT can be deposited on large areas, direct-conversion AMFPI of only $7.7 \times 7.7 \text{ cm}^2$ (512×512 pixels) have been demonstrated using polycrystalline CZT (poly-CZT). The CZT layer thickness varies in the range $200\text{--}500 \mu\text{m}$. Temporal lag and nonuniform response were noticeable in early CZT sensors. Large and nonuniform grain sizes are believed to be responsible for the temporal lag and nonuniform response of the sensor. Recent studies show that Cl doping makes a finer, more uniform grain structure [48.15]. The ionization energy $W_{\pm} \approx 5 \text{ eV}$, and the density of band gap Cd_{0.95}Zn_{0.05}Te is 5.8 g/cm^3 . The band gap energy E_g of Cd_{0.95}Zn_{0.05}Te is 1.7 eV and the resistivity of this material is $\approx 10^{11} \Omega \text{ cm}$ [48.32].

Typically CZT is first coated onto an inert substrate and then attached to the active matrix array. The advantage of this technique is that the electrical properties and hence imaging performances of the detector can be optimized without causing any thermal or chemical damage to the TFT array. The CZT film is deposited by the close-spaced sublimation method [48.33] onto an alumina (Al₂O₃) substrate coated with ITO, which forms the bias (top) electrode. A cross section of the poly-CZT detector is shown in Fig. 48.6. The several-

microns-thick ZnTe layer acts as a barrier to electron injection and hence suppresses dark current under the negative bias. This ZnTe layer is omitted for the positive bias. Conducting resin bumps on the pixel electrodes connect each pixel to the CZT layer.

Introduction of Zn into the CdTe lattice increases the band gap, decreases the conductivity and hence reduces the dark current. In CZT μ_h decreases with increasing Zn concentration whereas μ_e remains nearly constant. Again, addition of Zn to CdTe increases lattice defects and hence reduces carrier lifetimes. The poly-CZT has a lower density, which results in a lower quantum efficiency η than its single-crystal counterpart. Although (for a detector of given thickness) X-ray sensitivity in CZT detectors is lower than in CdTe detectors, the CZT detectors show a better signal-to-noise ratio and hence give a better DQE. The measured sensitivities are higher than a-Se.

The dark current of Cd_{0.95}Zn_{0.05}Te imagers increases almost linearly with F and is $\approx 70 \text{ pA/mm}^2$ at $F = 0.25 \text{ V}/\mu\text{m}$ [48.17], which makes it unsuitable for applications requiring long exposure times. The dark current would be expected to decrease with increasing Zn concentration due to the increased E_g . This has in fact, been demonstrated [48.15], i.e., the dark current in Cd_{0.92}Zn_{0.08}Te sensors is 40 pA/mm^2 at $F = 0.4 \text{ V}/\mu\text{m}$.

The mobility-lifetime products of both electrons and holes in poly-CZT are less than in single-crystal CZT. The values of $\mu\tau$ in single-crystal Cd_{0.9}Zn_{0.1}Te are in the range $10^{-4}\text{--}10^{-3} \text{ cm}^2/\text{V}$ (electrons) and $10^{-6}\text{--}10^{-5} \text{ cm}^2/\text{V}$ (holes) [48.34]. But the values of $\mu\tau$ in poly-Cd_{0.95}Zn_{0.05}Te are $\approx 2 \times 10^{-4} \text{ cm}^2/\text{V}$ (electrons), and $\approx 3 \times 10^{-6} \text{ cm}^2/\text{V}$ (holes) [48.35, 36]. Since $\mu_e\tau_e \gg \mu_h\tau_h$ in CZT, negative bias to the radiation-receiving electrode is the preferred choice for better sensitivity and temporal response.

The relative standard deviation of the sensitivity (standard deviation/average value) in the latest CZT AMFPI is $\approx 20\%$ [48.15]. The image-lag characteristics reported are $\approx 70\%$ first-frame lag, $\approx 20\%$ after three frames and 10% at 1 s in fluoroscopic mode (30 frames/s). In single-pulse radiographic mode the first-frame lag is less than 10% [48.15]. The longer image-lag characteristics of CZT sensors in fluoroscopic mode imply that it is not yet suitable for fluoroscopic applications. The presampling MTF of CZT detector is ≈ 0.3 ($150\text{-}\mu\text{m}$ pixel size) at the Nyquist frequency f_N (theoretical MTF, sinc function, is ≈ 0.64 at $f_N = 3.3 \text{ line pairs per mm}$), where the MTF of CsI imagers is less than 0.2 [48.17].

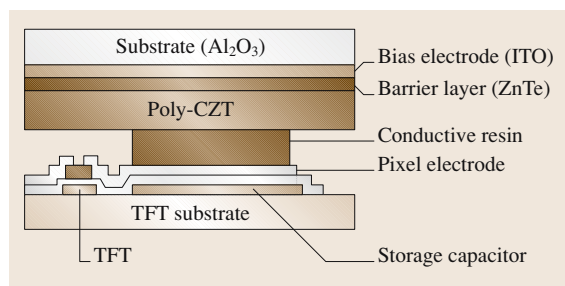


Fig. 48.6 Cross section of the polycrystalline CZT detector structure (after Tokuda et al. [48.15])

Lead Iodide (PbI₂)

PbI₂ polycrystalline photoconductive layers have been prepared using PVD at a substrate temperature of 200–230 °C. A deposition of several hundred angstroms of palladium (Pd) is used to form a top electrode. Grains are described as hexagonal platelets with the longest dimensions being 10 μm or less. The platelets grow perpendicular to the substrate, producing films that are less dense (3–5 g/cm³) than bulk crystalline material (6.2 g/cm³). There appears to be no technological barrier to preparing large-area layers, and direct-conversion AMFPI of 20×25 cm² size (1536×1920 pixels) have been demonstrated [48.37]. Coating thickness varies in the range 60–250 μm and prototype PbI₂ imagers have been used for radiographic imaging [48.37]. The band-gap energy $E_g = 2.3$ eV, and the ionization energy $W_{\pm} \approx 5$ eV. The resistivity of this material is in the range $10^{11} - 10^{12}$ Ω cm.

Lead iodide detectors have a very long image-lag decay time. The image lag depends on the exposure history. The image-lag characteristics reported are $\approx 75\%$ first-frame lag, $\approx 15\%$ after 3 s in fluoroscopic mode (15 frames/s), whereas in single-pulse radiographic mode the first-frame lag is less than 50% and drops below 1% within 1 s [48.37]. The long image-lag characteristics of PbI₂ in fluoroscopic mode imply that it is unsuitable for fluoroscopic applications.

The dark current of PbI₂ imagers increases sublinearly with the applied bias voltage. The dark current is in the range 10–50 pA/mm² at $F = 0.5$ V/μm, much higher than PVD HgI₂ detectors, making it

unsuitable for long-exposure-time applications. The presampling MTF is ≈ 0.35 (127 μm pixel size) at f_N (theoretical MTF, sinc function, is ≈ 0.64 at $f_N = 3.93$ lp/mm), where the MTF of CsI imagers is less than 0.2 [48.37]. The resolution of PbI₂ imagers is acceptable but slightly worse than that of HgI₂ imagers. Also, the X-ray sensitivity of PbI₂ imagers is lower than that of HgI₂ imagers. The pixel-to-pixel sensitivity variation in PbI₂ imagers is substantially lower.

The $\mu\tau$ product of holes and electrons in PVD PbI₂ are 1.8×10^{-6} cm²/V and 7×10^{-8} cm²/V, respectively [48.37]. μ_h in poly-PbI₂ is in the range 0.02–0.15 cm²/Vs whereas μ_h in single-crystal PbI₂ is 2 cm²/Vs [48.38]. This indicates that μ_h in poly-PbI₂ is controlled by shallow traps, probably introduced at the grain boundaries.

Lead Oxide (PbO)

The large-area deposition requirement is compatible with the use of polycrystalline PbO (poly-PbO) film as a photoconductor layer in AMFPI. Direct-conversion flat-panel X-ray imagers of 18×20 cm² (1080×960 pixels) from a poly-PbO with film thicknesses of ≈ 300 μm have been demonstrated [48.18]. One advantage of PbO over other X-ray photoconductors is the absence of heavy-element K-edges for the entire diagnostic energy range up to 88 keV, which suppresses additional noise and blurring due to the K-fluorescence. The ionization energy $W_{\pm} \approx 8$ eV and the density of poly-PbO is 9.6 g/cm³. The band-gap energy E_g of PbO

Table 48.1 Material properties

Photoconductor, state and preparation	E_g (eV)	W_{\pm} (eV)	Density (g/cm ³)	Resistivity (Ω cm)	Electrons μ_e (cm ² /Vs), $\mu_e \tau_e$ (cm ² /V)	Holes μ_h (cm ² /Vs) $\mu_h \tau_h$ (cm ² /V)
Stabilized a-Se, vacuum deposition	2.1–2.2	≈ 45 at 10 V/μm	4.3	$10^{14} - 10^{15}$	$\mu_e = 0.003 - 0.006$ $\mu_e \tau_e = 0.3 \times 10^{-6} - 10^{-5}$	$\mu_h = 0.12$ $\mu_h \tau_h = 10^{-6} - 6 \times 10^{-5}$
HgI ₂ , polycrystalline, PVD	2.1	5	6.3	$\approx 4 \times 10^{13}$	$\mu_e = 88$ $\mu_e \tau_e \approx 10^{-5} - 10^{-4}$	$\mu_h = 3 - 4$ $\mu_h \tau_h \approx 10^{-6}$
HgI ₂ , polycrystalline, SP	2.1	5	6.3	$\approx 4 \times 10^{13}$	$\mu_e \tau_e \approx 10^{-6} - 10^{-5}$	$\mu_h \tau_h \approx 10^{-7}$
Cd _{0.95} Zn _{0.05} Te, polycrystalline, vacuum deposition	1.7	5	5.8	$\approx 10^{11}$	$\mu_e \tau_e \approx 2 \times 10^{-4}$	$\mu_h \tau_h \approx 3 \times 10^{-6}$
PbI ₂ , polycrystalline, PVD	2.3	5	3–5	$10^{11} - 10^{12}$	$\mu_e \tau_e = 7 \times 10^{-8}$	$\mu_h = 0.02 - 0.15$ $\mu_h \tau_h \approx 2 \times 10^{-6}$
PbO, polycrystalline, vacuum deposition	1.9	8–20	9.6	$7 - 10 \times 10^{12}$	$\mu_e \tau_e \approx 5 \times 10^{-7}$?
TlBr, polycrystalline	2.7	6.5	7.5	$\approx 5 \times 10^9$ at 20 °C	?	$\mu_h \tau_h \approx 1.5 \times 10^{-6}$

Table 48.2 Imaging properties

Photoconductor, state and preparation	Typical operating F (V/ μm)	Dark current (pA/ mm^2)	Lag (fluoroscopic mode of operation)	Uniformity/sensitivity variation (standard deviation/average value)
Stabilized a-Se, single layer	≈ 10	< 10 up to $F = 20 \text{ V}/\mu\text{m}$	$< 2\%$ after 33 ms	Negligible
Stabilized a-Se, multilayer (PIN or NIP)	≈ 10	< 1 up to $F = 20 \text{ V}/\mu\text{m}$	$< 2\%$ after 33 ms	Negligible
HgI ₂ , polycrystalline, PVD	≈ 0.5	≈ 6 at $F = 0.5 \text{ V}/\mu\text{m}$	$\approx 7\%$ after 66 ms	$\approx 10\%$
HgI ₂ , polycrystalline, SP	≈ 1.0	≈ 8 at $F = 1.0 \text{ V}/\mu\text{m}$	$\approx 7\%$ after 66 ms	$\approx 10\%$
Cd _{0.95} Zn _{0.05} Te, polycrystalline	≈ 0.25	≈ 25 at $F = 0.25 \text{ V}/\mu\text{m}$	$\approx 70\%$ after 33 ms	$\approx 20\%$
PbI ₂ , polycrystalline, PVD	≈ 0.5	10–50 at $F = 0.5 \text{ V}/\mu\text{m}$	$\approx 75\%$ after 66 ms	?
PbO, polycrystalline	≈ 1.0	40 at 3 V/ μm field	3–8% after 1 s	?
TlBr, polycrystalline	≈ 1.0	?	?	?

is 1.9 eV and the resistivity of this material is in the range $(7\text{--}10)\times 10^{12} \Omega \text{ cm}$ [48.18].

Lead oxide photoconductive polycrystalline layers have been prepared by thermal evaporation in a vacuum chamber at a substrate temperature of $\approx 100^\circ\text{C}$. An evaporated layer of Al, Au or Pd was used to form a top electrode of thickness 100–200 nm [48.39]. The PbO layer consists of very thin platelets of a few microns thickness and has a density of $\approx 50\%$ of the single-crystal density. PbO slowly degrades if it is exposed to air under normal ambient temperature but a few hours exposure is acceptable. However, in the long term PbO reacts with water and CO₂ causing an increase in dark current and a decrease in X-ray sensitivity. Therefore, a polymer or a semiconductor (e.g., doped a-Se) passivation layer is used to prevent exposure to the atmosphere. This is usually placed between the photoconductor layer and the top metal electrode. The passivation layer avoids degradation of the PbO layer and reduces the dark current [48.39]. The dark current in PbO sensors is $\approx 40 \text{ pA}/\text{mm}^2$ at $F = 3 \text{ V}/\mu\text{m}$ [48.18]. $\mu_e\tau_e$ in poly-PbO is $\approx 5\times 10^{-7} \text{ cm}^2/\text{V}$ [48.18]. The lag signal in fluoroscopic mode is in the range 3–8% after 1 s. The presampling MTF of PbO detector is ≈ 0.5 (184 μm pixel size) at f_N (theoretical MTF, sinc function, is ≈ 0.64 at $f_N = 2.72 \text{ lp}/\text{mm}$) [48.18].

Thallium Bromide (TlBr)

Polycrystalline thallium bromide (poly-TlBr) has not yet been used in an AMFPI but has been used in a large-area (9-inch-diameter) direct-conversion detector called

an X-ray-sensitive electron-beam image tube (XEBIT). The operational principle of the XEBIT is similar to the standard light-sensitive vidicon that was utilized extensively in the commercial television industry. The XEBIT can replace an X-ray image intensifier coupled to a video camera using relay lenses with a single direct-conversion device [48.40]. The typical TlBr layer thickness is 300 μm .

The ionization energy $W_{\pm} \approx 6.5 \text{ eV}$ and the relative dielectric constant of TlBr is 33. The band-gap energy E_g of TlBr is 2.7 eV and its resistivity is $\approx 5\times 10^9 \Omega \text{ cm}$ under ambient conditions. At room temperature the dominant contribution to the dark current is ionic conductivity [48.40]. The ionic conductivity has an exponential dependence on temperature; the conductivity decreases by an order of magnitude for every 19 $^\circ\text{C}$ temperature decrease. Therefore, the dark current can be greatly decreased by Peltier cooling. $\mu_h\tau_h$ in TlBr is $\approx 1.5\times 10^{-6} \text{ cm}^2/\text{V}$ [48.40].

48.1.3 Summary and the Future

The material and imaging properties of potential photoconductors for X-ray image detectors are summarized in Tables 48.1 and 48.2. Stabilized a-Se is currently the best choice of photoconductor for clinical X-ray image detectors. The next closest competitor is the poly-HgI₂ imagers, which show excellent sensitivity, good resolution, and acceptable dark current, homogeneity and lag characteristics. However, the long-term stability of HgI₂ imagers has not been as thoroughly studied as

stabilized a-Se sensors. Both the dark current and the image-lag characteristics of CZT, PbI₂ and PbO detectors are worse than those of HgI₂ sensors. However, the X-ray detectors made with CZT photoconductive layers should be mechanically and chemically more stable than HgI₂-based detectors. The main drawback of a-Se detectors is its low conversion gain, which particularly affects the imaging sensor performance at low exposure. This can be overcome by utilizing the avalanche mul-

tiplication technique in the a-Se layer [48.41] and/or using on-pixel amplification [48.2]. However, further research is necessary to demonstrate basic operation as well as to examine the long-term stability of sensors utilizing these techniques. The main drawbacks of polycrystalline sensors are the image lag and the nonuniform response. Making smaller, finer and more uniform grain size in polycrystalline sensors may overcome these drawbacks.

48.2 Metrics of Detector Performance

X-ray sensitivity, resolution in terms of modulation transfer function (MTF), detective quantum efficiency (DQE), image lag and ghosting are often considered as the metrics of imaging performance. For most practical applications, the spatial-frequency-dependent (f -dependent) detective quantum efficiency, $DQE(f)$, is the appropriate metric of overall system performance and is unity at all f for an ideal detector. The detector performance depends critically on the photoconductor material properties such as the mobility, carrier trapping (both shallow and deep), EHP creation energy, X-ray attenuation and absorption coefficients. The material properties such as carrier mobility, EHP creation energy, X-ray attenuation and absorption coefficients in a well-defined photoconductor are almost constant, but the carrier lifetimes may vary from sample to sample. Shallow and deep trapping are particularly responsible for image lag and ghosting, respectively. The effects of charge-transport properties ($\mu\tau$) and the attenuation coefficient of photoconductor materials on the detector performance depends on L and F through the following normalized parameters,

$$\begin{aligned}\Delta &= \text{normalized attenuation depth} \\ & \quad (\text{attenuation depth/thickness}) \\ &= 1/(\alpha L), \\ x_e &= \text{normalized electron schubweg} \\ & \quad (\text{electron schubweg per unit thickness}) \\ &= \mu_e \tau_e F/L, \text{ and,} \\ x_h &= \text{normalized hole schubweg} \\ & \quad (\text{hole schubweg per unit thickness}) \\ &= \mu_h \tau_h F/L.\end{aligned}$$

where α is the linear attenuation coefficient of the photoconductor, $\mu_{e(h)}$ is the mobility and $\tau_{e(h)}$ is the deep-trapping time (lifetime) of electrons (holes).

Equivalently, x_e and x_h are the normalized carrier lifetimes (carrier lifetimes per unit transit time) for electrons and holes, respectively. The ranges of these normalized parameters for the three most promising photoconductive (a-Se, poly-HgI₂ and poly-CZT) X-ray image detectors are given in Table 48.3. The combined effects of charge-transport properties (mobility and carrier lifetime), operating conditions (F and E), photoconductor thickness, and the attenuation coefficient of the photoconductor material on the imaging characteristics (X-ray sensitivity, DQE, MTF and ghosting) are examined in the following sections. It must be emphasized that the photoconductor thickness L and the operating field F are as important to the overall performance of the detector as the material properties of the photoconductor itself, a point that will become apparent in the results presented in this chapter.

48.2.1 X-Ray Sensitivity

The X-ray sensitivity (S) of a photoconductive detector is defined as the collected charge per unit area per unit exposure of radiation and is considered an important performance measure for a superior image. High S permits the use of low detector radiation-exposure levels which also increases the dynamic range of the AMPFI. The selection of the X-ray photoconductor is highly influenced by the value of S .

The value of S can be considered to arise in terms of three controlling factors: firstly, the amount of radiation actually absorbed from the incident radiation that is useful for the generation of electron-hole pairs (EHPs), which is characterized by the quantum efficiency η of the detector and depends on the value of α of the photoconductor and L through $\eta = 1 - e^{-\alpha L}$, where the value of α is X-ray photon-energy-dependent; secondly, the generation of EHPs by X-ray interactions, which is char-

Table 48.3 The values of Δ , x_e and x_h for a-Se, poly-HgI₂ and poly-CZT detectors

Photoconductor	$\mu_e \tau_e$ (cm ² /V) $\mu_h \tau_h$ (cm ² /V)	F (V/ μ m)	E (keV)	L (mm)	x_e	x_h	Δ
Stabilized a-Se	$\mu_e \tau_e = 0.3 \times 10^{-6} - 10^{-5}$	≈ 10	20	0.2	1.5–50	5–300	0.24
	$\mu_h \tau_h = 10^{-6} - 6 \times 10^{-5}$		60	1.0	0.3–10	1–60	0.98
Poly-HgI ₂	$\mu_e \tau_e \approx 10^{-6} - 10^{-5}$	0.5–1	20	0.15	0.7–7	≈ 0.1	0.21
	$\mu_h \tau_h \approx 10^{-7}$		60	0.3	0.35–3.5	≈ 0.05	0.85
Poly-Cd _{0.95} Zn _{0.05} Te	$\mu_e \tau_e \approx 2 \times 10^{-4}$	≈ 0.25	20	0.3	≈ 17	≈ 0.25	0.26
	$\mu_h \tau_h \approx 3 \times 10^{-6}$		60	0.3	≈ 17	≈ 0.25	0.89

acterized by the value of W_{\pm} of the photoconductor and the average absorbed energy (E_{ab}) per attenuated X-ray photon of energy E , where W_{\pm} depends on the material properties of the photoconductor, and E_{ab} depends on the incident X-ray photon energy [48.42] and the material properties; and thirdly, how much of the X-ray generated charge is actually collected in the external circuit. This is characterized by the charge-carrier drift mobilities (μ) and lifetimes (τ), the applied F and L .

The S of an X-ray image detector can be normalized with respect to the maximum sensitivity (S_0) that would arise if all the incident radiation were absorbed and all the liberated carriers were collected. Neglecting secondary photon interactions, the expression for S_0 is [48.22]

$$S_0 = \left(\frac{5.45 \times 10^{13} e}{(\alpha_{air}/\rho_{air}) W_{\pm}} \right) \left(\frac{\alpha_{en}}{\alpha} \right), \quad (48.3)$$

where e is the elementary charge, α_{en} is the energy absorption coefficient of the photoconductor, while α_{air} and ρ_{air} are the energy absorption coefficient of air and its density. If W_{\pm} is expressed in eV, α_{air}/ρ_{air} is in cm²/g and exposure in (48.3) is in Roentgens, then S_0 is in units of C/cm²R. Thus S_0 is a constant that depends on E and the material properties of the photoconductor, since W_{\pm} is a material property which can usually be taken as constant for a given material. For those materials (e.g., a-Se) that have a significant F - and/or E -dependent W_{\pm} , then S_0 depends on F and/or E .

The quantity $s = S/S_0$ takes into account the X-ray absorption and charge-transport effects and is called the charge-collection and absorption-limited normalized sensitivity. It should be emphasized that s is a quantity that is determined by the X-ray absorption profile, photoconductor thickness and the charge-collection efficiency. The s of an X-ray image detector considering small signal operation, a constant μ and a single deep-trapping time (lifetime) τ for each type of carrier (holes and electrons) and neglecting carrier diffusion is given

by [48.43],

$$\begin{aligned} S/S_0 &= x_h \left[\left(1 - e^{-\frac{1}{\Delta}} \right) + \frac{1}{\Delta/x_h - 1} \left(e^{-\frac{1}{x_h}} - e^{-\frac{1}{\Delta}} \right) \right] \\ &\quad + x_e \left[\left(1 - e^{-\frac{1}{\Delta}} \right) - \frac{1}{\Delta/x_e + 1} \left(1 - e^{-\frac{1}{\Delta} - \frac{1}{x_e}} \right) \right] \\ &= s_h(x_h, \Delta) + s_e(x_e, \Delta) = s(x_h, x_e, \Delta) \end{aligned} \quad (48.4)$$

where subscripts h and e refer to holes and electrons respectively.

The two square brackets on the right-hand side of the normalized sensitivity s expression (48.4) represent the relative contributions of hole and electron transport to the overall sensitivity for a given Δ . It is assumed in (48.4) that the radiation receiving side of the detector is biased positively. If the bias polarity is reversed, then x_e and x_h must be interchanged. The expression in (48.4) applies for incident radiation that is monoenergetic and has to be appropriately integrated over the radiation spectrum of a practical polyenergetic X-ray source by considering the X-ray photon-energy-dependent terms W_{\pm} , α and α_{en} . Equation 48.4 applies to an isolated photoconductor sandwiched between two large-area parallel-plate electrodes (small pixel effects are excluded) [48.44] and operating under a constant E (small signal case). An excellent fit of (48.4) to experimental data on poly-HgI₂ is given in [48.30].

The value of s is always less than unity since S for a photoconductor of finite thickness in which carrier collection is not perfect is always less than S_0 . Note that $s(x_h, x_e, \Delta) = s_h + s_e = 1$ when all the incident radiation is absorbed and all the charges are collected, i.e., $x_h, x_e \gg 1$ and $\Delta \ll 1$. The sensitivity is then simply S_0 and is controlled by W_{\pm} .

The sensitivity is mainly controlled by the charges that have the same polarity as the bias on the radiation-receiving electrode: holes for positive bias and electrons for negative bias [48.30, 43]. The extent of the disparity between s_h and s_e depends on Δ . The disparity

Table 48.4 X-ray sensitivity of a-Se, poly-HgI₂ and poly-CZT detectors using the normalized parameters from Table 48.3

Photoconductor	E (keV)	S_0 ($\mu\text{C}/\text{cm}^2\text{R}$)	$s = S/S_0$		S ($\mu\text{C}/\text{cm}^2\text{R}$)	
			Positive bias	Negative bias	Positive bias	Negative bias
Stabilized a-Se	20	0.244	0.9–0.98	0.8–0.98	0.22–0.24	0.2–0.24
	60	5.37	0.39–0.64	0.35–0.62	2.1–3.38	1.88–3.35
Poly-HgI ₂	20	2.75	0.25–0.29	0.53–0.81	0.7–0.81	1.46–2.24
	60	38.54	0.15–0.3	0.21–0.4	6.76–11.21	8.18–15.6
Poly-Cd _{0.95} Zn _{0.05} Te	20	3	≈ 0.456	≈ 0.85	≈ 1.37	≈ 2.55
	60	35.87	≈ 0.41	≈ 0.51	≈ 14.69	≈ 18.12

is stronger for lower Δ , which can be understood by noting that electron and hole generation do not occur uniformly throughout the thickness of the sample but rather it is more effective closer to the radiation receiving electrode. An advantage of a-Se detectors is that both x_e and x_h are much greater than one, which is not the case for other photoconductors, as shown in Table 48.3. Therefore, charge collection in a-Se detectors is close to unity and the normalized sensitivity is controlled by the quantum efficiency of the detector. The S and s of different photoconductive detectors using the normalized parameters from Table 48.3 are given in Table 48.4. The values of W_{\pm} for a-Se were taken from the work of Blevis et al. [48.20]. At $F = 10 \text{ V}/\mu\text{m}$, $W_{\pm} = 42.5$ and 46 eV for X-ray photon energies of 60 and 20 keV, respectively. The maximum X-ray sensitivity S_0 of chest

radiographic detectors ($E = 60 \text{ eV}$) is much higher than of mammographic detectors ($E = 20 \text{ eV}$) because of the lower $\alpha_{\text{air}}/\rho_{\text{air}}$ values in (48.3) at higher X-ray photon energies [48.42]. The value of S_0 in a-Se detectors is much lower than in poly-HgI₂ and poly-CZT detectors because of the higher value of W_{\pm} in a-Se.

48.2.2 Detective Quantum Efficiency

DQE measures the ability of the detector to transfer signal relative to noise from its input to its output. The random nature of image quanta gives rise to random fluctuations in image signals, which contributes to image formation and hence creates random noises. The scattering of image quanta gives rise to image blurring, which is quantified by the $\text{MTF}(f)$. Images are partially degraded by various sources of statistical fluctuations that arise along the imaging chain. The relative increase in image noise due to an imaging system as a function of spatial frequency f is expressed quantitatively by $\text{DQE}(f)$ which represents the signal-to-noise transfer efficiency for different frequencies of information in an image. $\text{DQE}(f)$ is defined as

$$\text{DQE}(f) = \frac{\text{SNR}_{\text{out}}^2(f)}{\text{SNR}_{\text{in}}^2(f)}, \quad (48.5)$$

where SNR_{in} and SNR_{out} are the signal-to-noise ratios (SNR) at the input and output stages of the image detector, respectively. $\text{DQE}(f)$ is unity for an ideal detector. For simplicity, we are often interested in measuring $\text{DQE}(f = 0)$ of an imaging detector since it represents the signal quality degradation due to the signal and noise transfer characteristics of the system without considering signal spreading.

The random nature of charge-carrier trapping (i. e. incomplete charge collection) in the photoconductor layer creates fluctuations in the collected charge and hence creates additional noise. Thus carrier trapping degrades signal-to-noise performance of the image and reduces the DQE. Recently, Kabir and Kasap [48.45]

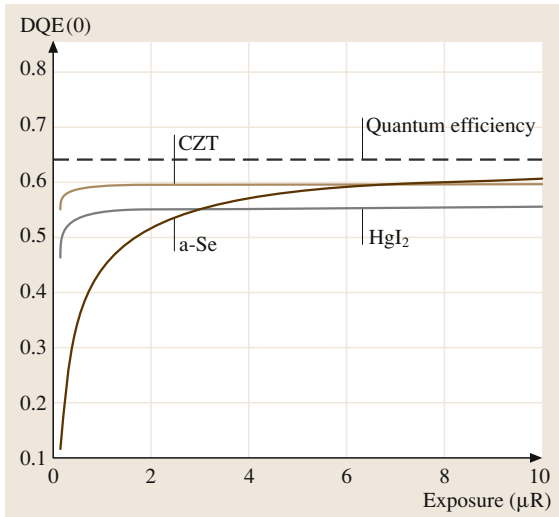


Fig. 48.7 $\text{DQE}(0)$ versus X-ray exposure for a-Se, poly-HgI₂, and poly-CZT detectors and for a 60-keV monoenergetic X-ray beam. The electronic noise is 2000 e per pixel. It is assumed that $F = 10 \text{ V}/\mu\text{m}$ for a-Se, $0.5 \text{ V}/\mu\text{m}$ for HgI₂ and $0.25 \text{ V}/\mu\text{m}$ for CZT

have examined the effects of charge-carrier trapping on the $DQE(0)$ of an a-Se detector by considering depth-dependent conversion gain and depth-dependent charge-collection efficiency in the cascaded-linear-system model [48.46, 47]. In this section, the $DQE(0)$ model of *Kabir and Kasap* [48.45] is applied to potential photoconductive detectors such as a-Se, poly-HgI₂ and poly-Cd_{0.95}Zn_{0.05}Te detectors for fluoroscopic applications to study and compare their $DQE(0)$ performance.

Figure 48.7 shows $DQE(0)$ as a function of X-ray exposure for a-Se, HgI₂, and CZT detectors for a 60-keV X-ray beam. The X-ray exposure (X) is varied from 0.1 μ R to 10 μ R, which is the range of X-ray exposure for fluoroscopic applications. We assume that the pixel area, $A = 200 \mu\text{m} \times 200 \mu\text{m}$, and the effective fill factor is 1.0 for all types of photoconductors. The average E is 60 keV and the additive electronic noise (N_e) is assumed to be 2000 electrons per pixel. The following transport and operating parameters are used in Fig. 48.7: for a-Se detectors, $L = 1000 \mu\text{m}$, $F = 10 \text{ V}/\mu\text{m}$, $W_{\pm} \approx 43 \text{ eV}$, $\mu_e \tau_e = 10^{-6} \text{ cm}^2/\text{V}$ and $\mu_h \tau_h = 10^{-5} \text{ cm}^2/\text{V}$; for HgI₂ detectors, $L = 260 \mu\text{m}$, $F = 0.5 \text{ V}/\mu\text{m}$, $W_{\pm} = 5 \text{ eV}$, $\mu_e \tau_e = 6 \times 10^{-6} \text{ cm}^2/\text{V}$ and $\mu_h \tau_h = 10^{-7} \text{ cm}^2/\text{V}$; and for CZT detectors, $L = 270 \mu\text{m}$, $F = 0.25 \text{ V}/\mu\text{m}$, $W_{\pm} = 5 \text{ eV}$, $\mu_e \tau_e = 2 \times 10^{-4} \text{ cm}^2/\text{V}$ and $\mu_h \tau_h = 3 \times 10^{-6} \text{ cm}^2/\text{V}$. The radiation-receiving electrode is biased positively for a-Se detectors and negatively for HgI₂ and CZT detectors. $\Delta \approx 0.98$ is for all photoconductors. $\eta = 1 - \exp(-1/\Delta) \approx 0.64$ is the maximum achievable $DQE(0)$ if all the liberated charges are collected.

The DQE for the CZT detector is relatively unchanged over the whole exposure range due to a large conversion gain (low W_{\pm}) and high charge-collection efficiency (good transport properties). The DQE for a-Se detectors is small at low exposures because of its relatively low conversion gain and is controlled by the added electronic noise. As the X-ray exposure increases, each pixel receives more photons. The relative contribution of electronic noise to the total noise becomes less important and the DQE increases. The DQE for HgI₂ detectors is relatively small even at higher exposures because of its low charge-collection efficiency, which gives rise to considerable gain-fluctuation noise. Therefore, both high conversion gain and high charge-collection efficiency are required to improve the DQE performance of an X-ray image detector [48.48]. The conversion gain depends on W_{\pm} , which is a material property of the photoconductor. The charge-collection efficiency can be improved by increasing F and improving the $\mu\tau$ products of the carriers. However, increasing

F also increases the dark current dramatically in both HgI₂ and CdZnTe detectors. Thus there is a practical limitation on F . An F as high as $20 \text{ V}/\mu\text{m}$ is achievable in a-Se detectors while keeping the dark current within an acceptable level for X-ray imaging [48.26]. The charge-collection efficiency of a-Se detectors is relatively high because of the high F needed to obtain a reasonable W_{\pm} .

48.2.3 Modulation Transfer Function (MTF)

Resolution or resolving power is the ability to record separate images of small objects that are placed very closely together. The overall resolution of a system can be expressed as a convolution of the component resolutions. However, the spatial resolution of an imaging device or a system can also be described in terms of the MTF , which is the relative response of the system as a function of spatial frequency. The MTF of an imaging system can be described as a cascade of several stages where the overall MTF is simply the product of the $MTFs$ of all the individual stages. The $MTF(f)$ is a much more convenient descriptor of spatial response since the resolving power from multiplication is much more easily comprehended than convolution. The overall MTF (or presampling MTF) of an image detector can be expressed as,

$$MTF(f) = MTF_m(f) \times MTF_a(f), \quad (48.6)$$

where $MTF_m(f)$ is the modulation transfer function of the detector material and $MTF_a(f)$ is the modulation transfer function associated with the aperture function of the pixel electrodes. $MTF_a(f)$ arises due to averaging of the signal over a pixel area. If the aperture is square with dimension a , then, $MTF_a(f)$ will be of the form $\text{sinc}(af)$. The aperture MTF describes how spatial frequencies are passed through the detector elements.

The spatial resolution in direct-conversion AMFPI is high and closer to the aperture function, as compared to phosphor-based AMFPI. Based on current understanding, charge-carrier trapping and reabsorption of K-fluorescent X-ray photons are the two dominant mechanisms responsible for the loss of resolution [48.48] in direct-conversion AMFPI. Some of the K-fluorescent X-ray photons may be reabsorbed at different points within the detector volume from the primary X-ray-photon interaction point. This creates a lateral spreading of signal and a loss of resolution. The loss of resolution due to fluorescence reabsorption is maximum (although not very substantial [48.48]) just above the K-edge of the photoconductor. This effect can

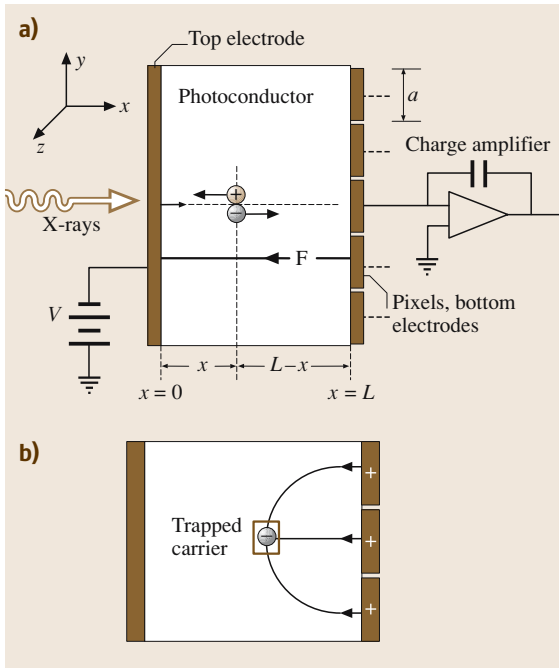


Fig. 48.8 (a) A cross section of a direct-conversion pixelated X-ray image detector. (b) Trapped carriers in the photoconductor induce charges not only on the central pixel electrode but also on neighboring pixel electrodes, spreading the information and hence reducing spatial resolution

be ignored when: (i) the incident X-ray photon energy is lower than the K-edge of the photoconductor, or (ii) the mean energy of the X-ray beam and the K-edge occur at widely different energies [48.49]. The charge-carrier trapping has a significant effect on the resolution of these direct-conversion X-ray image detectors [48.36], which will be discussed below.

The direct-conversion AMFPI geometry consists of a photoconductor layer sandwiched between two electrodes; the electrode at one side is a continuous metal plate and the electrode on the other side of the photoconductor is segmented into an array of individual square pixels of size $a \times a$, as shown in Fig. 48.8a. The geometric pixel aperture width in a flat-panel detector is smaller than the pixel pitch (center-to-center spacing between two pixels). However, it has been shown that the effective fill factor (the effective fraction of pixel area used for image-charge collection) of a photoconductive flat-panel detector is close to unity [48.28, 51]. Therefore, the effective pixel aperture width a is virtually identical to the pixel pitch. Some of the X-ray-generated carriers are captured by deep traps in the bulk of the photo-

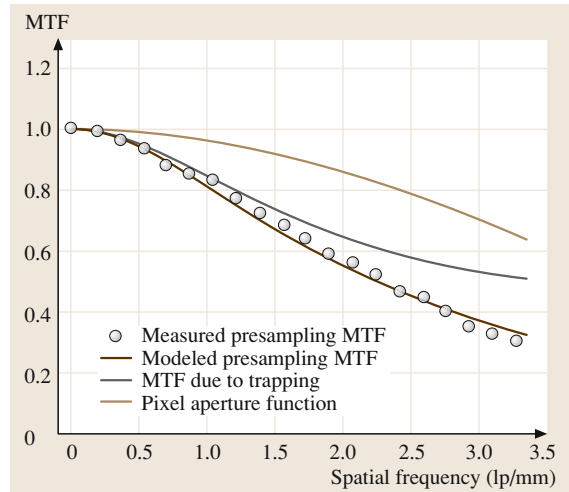


Fig. 48.9 Measured presampling MTF of a polycrystalline CdZnTe detector in comparison with modeled results that included blurring due to charge-carrier trapping in the bulk of the photoconductor. The detector thickness is 300 μm and the pixel pitch is 150 μm . (after Fig. 10 of [48.50])

conductor during their drift across the photoconductor layer. Suppose that a carrier is trapped in the photoconductor above a particular (central) pixel electrode of a pixelated image sensor. This trapped carrier induces charges not only on the central pixel electrode but also on neighboring pixel electrodes, as shown in Fig. 48.8b, and consequently there is a lateral spread of information and hence a loss of image resolution.

Kabir and Kasap [48.36] have developed an analytical expression for the MTF due to distributed carrier trapping in the bulk of the photoconductor and examined the effect of charge-carrier trapping on the resolution of direct-conversion AMFPI. The effect of trapping on MTF increases with decreasing normalized carrier lifetime (i.e., normalized schubweg). Trapping of the carriers that move towards the pixel electrodes degrades the MTF performance, whereas trapping of the other type of carriers, which move away from the pixels, improves the sharpness of the X-ray image.

Figure 48.9 shows the experimental MTF and theoretical fit of a 300- μm -thick positively biased CZT detector exposed to an 80-kVp X-ray beam with 26-mm Al filtration. The experimental data have been extracted from Fig. 10 of [48.50]. The operating $F = 0.25 \text{ V}/\mu\text{m}$ and pixel pitch is 150 μm . The Nyquist frequency is 3.3 lp/mm. The dotted line shows the MTF due to the bulk carrier trapping only. The bulk carrier trapping has a significant effect on the

MTF of the detector. As is apparent from Fig. 48.9, there is very good agreement between the model and the experimental data. The best-fit $\mu\tau$ products of electrons and holes $\mu_e\tau_e = 2.4 \times 10^{-4} \text{ cm}^2/\text{V}$ and $\mu_h\tau_h = 3.2 \times 10^{-6} \text{ cm}^2/\text{V}$, which are very close to the

$\mu\tau$ values reported previously [48.32, 35]. Although the charge-carrier-trapping-limited MTF model has been applied to the CZT sensors, the model can also be applied to other photoconductive (e.g., a-Se and HgI₂) panel X-ray image detectors.

48.3 Conclusion

The principles of operation of a direct-conversion AMFPI for medical applications have been briefly discussed. The charge-transport and imaging properties of some of the potential photoconductors have been critically discussed and compared with the properties of an ideal photoconductor for X-ray image detectors. The various imaging characteristics of photoconductor-based AMFPIs such as the sensitivity (*S*), detective quantum efficiency (DQE), and resolution in terms of the modulation transfer function (MTF) have also been examined. These characteristics depend critically not only on the photoconductor's charge-transport properties but also on the detector structure, i.e., the size of the pixel and the thickness of the photoconductor. It has been shown that the detector structure in terms

of the photoconductor thickness and the pixel size is just as important to the overall performance of the detector as the material properties of the photoconductor itself.

Recent experiments on a-Se detectors indicate that the X-ray-sensitivity of this detector decreases in subsequent exposures. The study of ghosting mechanisms in a-Se-based AMFPI is currently considered as an active research area since the origins of ghosting have not been fully resolved [48.52]. We expect that the ghosting phenomenon may also be present in other photoconductive (e.g., HgI₂, CZT, PbO and PbI₂) detectors, although it has not yet been measured. Long-term stability, and X-ray-induced effects and phenomena are also current research topics.

References

- 48.1 J. A. Rowlands, J. Yorkston: *Handbook of Medical Imaging*, Vol. 1, ed. by J. Beutel, H. L. Kundel, R. L. Van Metter (SPIE, Washington 2000) Chap. 4 and references therein for the various flat-panel X-ray image sensors
- 48.2 K. S. Karim, A. Nathan, J. A. Rowlands, S. O. Kasap: IEE Proc.-CDS **150**, 267 (2003)
- 48.3 V. N. Cooper III, T. Oshiro, C. H. Cagnon, L. W. Bassett, T. M. McLeod-Stockmann, N. V. Bezrukiy: Med. Phys. **30**, 2614 (2003)
- 48.4 D. C. Hunt, O. Tousignant, J. A. Rowlands: Med. Phys. **31**, 1166 (2004)
- 48.5 L. E. Antonuk, J. M. Boudry, Y. El-Mohri, W. Huang, J. H. Siewerdsen, J. Yorkston, R. A. Street: Proc. SPIE **2432**, 216 (1995)
- 48.6 J. Chabbal, C. Chaussat, T. Ducourant, L. Fritsch, J. Michailos, V. Spinnler, G. Vieux, M. Arques, G. Halm, M. Hoheisel, H. Horbaschek, R. Schulz, M. Spahn: Proc. SPIE **2708**, 499 (1996)
- 48.7 J. A. Rowlands, S. O. Kasap: Phys. Today **50**, 24 (1997)
- 48.8 O. Tousignant, M. Choquette, Y. Demers, L. Laperrière, J. Leboeuf, M. Honda, M. Nishiki, A. Takahashi, A. Tsukamoto: Proc. SPIE **4682**, 503 (2002)
- 48.9 S. O. Kasap, J. A. Rowlands: *Optoelectronics and Photonics: Principles and Practices* (Prentice-Hall, Upper Saddle River, New Jersey 2001)
- 48.10 S. O. Kasap, J. A. Rowlands: IEE Proc.-CDS **149**, 85 (2002)
- 48.11 S. O. Kasap: *Handbook of Imaging Materials*, ed. by A. S. Diamond (Marcel Dekker, New York 1991) Chap. 9
- 48.12 S. O. Kasap, J. A. Rowlands: Proc. IEEE **90**, 591 (2002)
- 48.13 F. Greuter, G. Blatter: Semicond. Sci. Technol. **5**, 111 (1990)
- 48.14 A. Zuck, M. Schieber, O. Khakhan, Z. Burshtein: IEEE Trans. Nuclear Sci. **50**, 991 (2003)
- 48.15 S. Tokuda, H. Kishihara, S. Adachi, T. Sato: Proc. SPIE **5030**, 861 (2003)
- 48.16 R. A. Street, S. E. Ready, K. Van Schuylenbergh, J. Ho, J. B. Boyec, P. Nylen, K. Shah, L. Melekhov, H. Hermon: J. Appl. Phys. **91**, 3345 (2002)
- 48.17 S. Tokuda, H. Kishihara, S. Adachi, T. Sato: J. Mater. Sci. Mater. Electron. **15**, 1 (2004)
- 48.18 M. Simon, R. A. Ford, A. R. Franklin, S. P. Grabowski, B. Mensor, G. Much, A. Nascetti, M. Overdick, M. J. Powell, D. U. Wiechert: Proc. SPIE **5368**, 188 (2004)

- 48.19 S. O. Kasap, K. V. Koughia, B. Fogal, G. Belev, R. E. Johanson: *Semiconductor* **14**, 816 (2003)
- 48.20 I. M. Blevis, D. C. Hunt, J. A. Rowlands: *J. Appl. Phys.* **85**, 7958 (1999)
- 48.21 M. F. Stone, W. Zhao, B. V. Jacak, P. O'Conner, B. Yu, P. Rehak: *Med. Phys.* **29**, 319 (2002)
- 48.22 S. O. Kasap: *J. Phys. D: Appl. Phys.* **33**, 2853 (2000)
- 48.23 W. Que, J. A. Rowlands: *Phys. Rev. B* **51**, 10 500 (1995)
- 48.24 E. Fourkal, M. Lachaine, B. G. Fallone: *Phys. Rev. B* **63**, 195 204 (2001)
- 48.25 D. Mah, J. A. Rowlands, J. A. Rawlinson: *Med. Phys.* **25**, 444 (1998)
- 48.26 B. Polischuk, Z. Shukri, A. Legros, H. Rougeot: *Proc. SPIE* **3336**, 494 (1998)
- 48.27 M. Choquette, Y. Demers, Z. Shukri, O. Tousignant, K. Aoki, M. Honda, A. Takahashi, A. Tsukamoto: *Proc. SPIE* **4320**, 501 (2001)
- 48.28 D. C. Hunt, O. Tousignant, Y. Demers, L. Laperrière, J. A. Rowlands: *Proc. SPIE* **5030**, 226 (2003)
- 48.29 M. Schieber, A. Zuck, M. Braiman, J. Nissenbaum, R. Turchetta, W. Dulinski, D. Husson, J. L. Riester: *IEEE Trans. Nucl. Sci.* **NS-44**, 2571 (1997)
- 48.30 M. Z. Kabir, S. O. Kasap: *Appl. Phys. Lett.* **80**, 1664 (2002)
- 48.31 R. C. Whited, L. Van den Berg: *IEEE Trans. Nucl. Sci.* **NS-24**, 165 (1977)
- 48.32 Y. Eisen, A. Shor: *Mater. Res. Soc. Symp. Proc.* **487**, 129 (1997)
- 48.33 T. L. Chu, S. S. Chu, C. Ferekides, C. Q. Wu, J. Britt, C. Wang: *J. Appl. Phys.* **70**, 7608 (1991)
- 48.34 A. Ruzin, Y. Nemirovsky: *J. Appl. Phys.* **82**, 4166 (1997)
- 48.35 J. G. Mainprize, N. L. Ford, S. Yin, E. E. Gordon, W. J. Hamilton, T. O. Tümer, M. J. Yaffe: *Med. Phys.* **29**, 2767 (2002)
- 48.36 M. Z. Kabir, S. O. Kasap: *J. Phys. D: Appl. Phys.* **36**, 2352 (2003)
- 48.37 G. Zentai, L. Partain, R. Pavlyuchkova, C. Proano, G. Virshup, L. Melekhov, A. Zuck, B. N. Breen, O. Dagan, A. Vilensky, M. Schieber, H. Gilboa, P. Bennet, K. Shah, Y. Dmitriev, J. Thomas: *Proc. SPIE* **5030**, 77 (2003)
- 48.38 R. A. Street, S. E. Ready, F. Lemmi, K. S. Shah, P. Bennett, Y. Dmitriyev: *J. Appl. Phys.* **86**, 2660 (1999)
- 48.39 A. Brauers, U. Schiebel: *US Patent* (1998) No. 5729 021
- 48.40 D. R. Ouimette, S. Nudelman, R. Aikens: *Proc. SPIE* **3336**, 470 (1998)
- 48.41 D. C. Hunt, S. S. Kirby, J. A. Rowlands: *Med. Phys.* **29**, 2464 (2002)
- 48.42 J. M. Boone: *Handbook of Medical Imaging*, Vol.1, ed. by J. Beutel, H. L. Kundel, R. L. Van Metter (SPIE, Washington 2000) Chap.1 and references therein
- 48.43 M. Z. Kabir, S. O. Kasap: *J. Vac. Sci. Tech. A* **20**, 1082 (2002)
- 48.44 M. Z. Kabir, S. O. Kasap: *J. Vac. Sci. Tech. A* **22**, 975 (2004)
- 48.45 M. Z. Kabir, S. O. Kasap: *J. Phys. D: Appl. Phys.* **35**, 2735 (2002)
- 48.46 M. Rabbani, R. Shaw, R. L. Van Metter: *J. Opt. Soc. Am. A* **4**, 895 (1987)
- 48.47 J. G. Mainprize, D. C. Hunt, M. J. Yaffe: *Med. Phys.* **29**, 976 (2002)
- 48.48 M. Z. Kabir, S. O. Kasap: *IEE Proc.-CDS* **150**, 258 (2003)
- 48.49 W. Que, J. A. Rowlands: *Med. Phys.* **22**, 365 (1995)
- 48.50 S. Tokuda, H. Kishihara, S. Adachi, T. Sato, Y. Izumi, O. Teranuma, Y. Yamane, S. Yamada: *Proc. SPIE* **4682**, 30 (2002)
- 48.51 G. Pang, W. Zhao, J. A. Rowlands: *Med. Phys.* **25**, 1636 (1998)
- 48.52 M. Z. Kabir, M. Yunus, S. O. Kasap, O. Tousignant, H. Mani, P. Gauthier: *Proc. SPIE* **5745**, 223 (2005)

49. Phase-Change Optical Recording

The present chapter takes the reader from the basics of phase-change recording up to the latest achievements in the field. It discusses in detail the specific features of Te-based compounds that made them the best materials for phase-change data storage. It is demonstrated that the essence of phase-change recording does not consist of simple disordering of the medium through melting and subsequent quenching as previously believed but is due to a switch of Ge atoms between octahedral and tetrahedral symmetry positions within the Te face-centered cubic lattice. It is this nature of the transition that makes the Te-based media fast and stable. The chapter is concluded by the introduction of a concept of the super-resolution

49.1	Digital Versatile Discs (DVDs)	1140
49.1.1	Questions Pertaining to DVD Recording	1140
49.1.2	Local Structure and its Changes During the Phase Transition	1140
49.1.3	Related Issues.....	1143
49.2	Super-RENS Discs	1144
49.3	In Lieu of Conclusion	1145
	References	1145

near-field structure (super-RENS) disc that allows the reduction of the bit size well below the diffraction limit and makes 100 GB/disc storage a reality.

The idea of phase-change recording dates back to the mid 1960s when *Ovshinsky* [49.1] first suggested using differences in electrical and optical properties between amorphous and crystalline phases of Te-based chalcogenides for data storage. The idea of phase-change recording is quite simple.

When a melt is cooled down slowly, such that the structure always remains in thermal equilibrium, upon

reaching the crystallization temperature the material crystallizes, i. e. is transformed into a solid state with a well-defined periodic structure. If, on the other hand, the cooling rate is fast, then at a certain temperature the viscosity of the liquid increases to a degree such that the structure can no longer relax following the changes in temperature; one obtains a supercooled liquid and then a glass. In contrast to the crystallization temperature, the glass-transition temperature is not well defined. A range of temperatures exist and the particular temperature of the glass transition depends on the cooling rate.

Once in the solid state, the glass, if kept at a temperature close to the glass-transition temperature, crystallizes. On the other hand, rapid heating of the crystalline material to a temperature above the melting point and subsequent rapid cooling (quenching) can produce a glassy state. This glass-formation diagram is demonstrated by Fig. 49.1.

The material can also be heated by light. Exposure of an amorphous material to a laser light that heats it above the glass-transition temperature results in crystallization while short and intense laser pulses melt the material and—provided the heat-dissipation rate is fast enough—an amorphous recorded bit is formed.

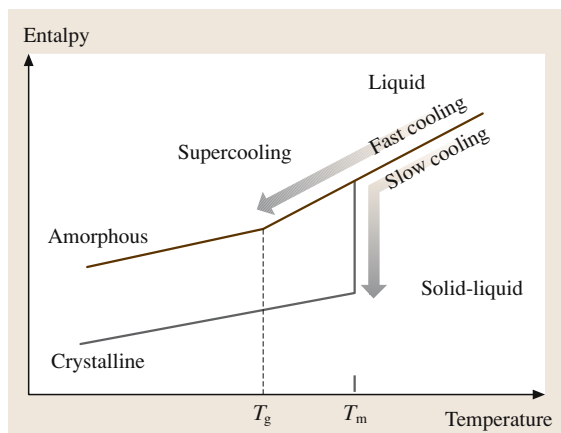


Fig. 49.1 A glass-formation diagram

49.1 Digital Versatile Discs (DVDs)

The latest commercial implementation of phase-change recording are rewritable digital versatile disks (DVDs) [49.2]. Various materials have been tried for potential use in phase-change memories and the best in terms of speed and stability are ternary Ge–Sb–Te (primarily $\text{Ge}_2\text{Sb}_2\text{Te}_5$, hereafter referred to as GST) and quaternary Ag–In–Sb–Te (AIST) alloys. The former are used in DVD random access memory (DVD-RAM) and the latter in DVD rewritable (DVD-RW) disks. While there are some differences in the device structure and recording strategies, both these kinds of rewritable DVDs are capable of storing up to 4.7 Gbits of information.

A basic disk structure consists of four layers (Fig. 49.2). The recording layer (GST or AIST) is sandwiched between Zn–SiO₂ layers and a metallic Al–Cr layer serves as a reflection layer and also as a heat sink to ensure the rapid cooling necessary for the formation of the amorphous phase.

49.1.1 Questions Pertaining to DVD Recording

While the phenomenology of phase-change recording is simple, the realistic nanoscale mechanism of the structural transformation has remained unclear for almost 30 years since the idea of phase-change recording was first suggested.

In particular, there are several fundamental questions that need to be answered. Interestingly, some of them have never been asked previously.

1. If the process consists of simple crystallization and melting there should be a wide variety of suitable composition. Why are GST and AIST the best materials?
2. Commercial DVD-RAM is stable for over 1 000 000 cycles. What makes the switching so reproducible and the media so stable?
3. GST crystallizes within 30 ns and amorphous bits can be recorded with femtosecond pulses. Why are the structural transformations so fast? Structure relaxation involving ordering of the amorphous material is likely to proceed on a longer time scale.
4. The crystallization process is characterized by an activation energy of 2–3 eV. Should the crystallization be a solid-state process, one would expect the activation energy to be on the order of the energy gap, i. e. ≈ 0.3 eV. The observed activation energy is rather

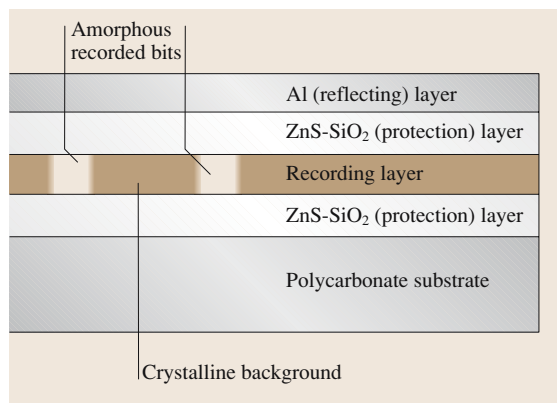


Fig. 49.2 A schematic representation of a DVD structure

reminiscent of a molecular-type excitation. Why is the activation energy so high?

5. In typical semiconductors (and in chalcogenides in particular), the amorphous phase is characterized by a smaller band gap than the crystalline phase. This is typically due to the fact that the bonds in the amorphous materials are weaker and hence the splitting between the bonding and antibonding states, which form the valence and conduction bands, respectively, is smaller. In phase-change materials the situation is reversed, namely, it is the crystalline phase that possesses a narrower band gap and higher reflectivity than the amorphous phase. What is the reason for this abnormality?
6. And, finally, what is the role of light? Although it has been tacitly assumed that light is simply heating the material, a very high concentration of nonequilibrium carriers created by intense light may—and most likely does—play a role.

In order to develop novel optical media insightfully, one has to answer the above questions. These, in turn, require detailed knowledge of the structure of the material in both the crystalline and amorphous phases as well as the atomic mechanism of the structural changes upon the phase transition.

49.1.2 Local Structure and its Changes During the Phase Transition

An ideal tool to investigate the local structure of a material and its changes on the atomic scale independent of the state of the material (crystalline or

amorphous) is X-ray absorption fine-structure (XAFS) spectroscopy.

Extended X-ray absorption fine structure (EXAFS) allows one to obtain information about the local structure around selected chemical species, such as the average coordination number, the bond lengths, the chemical nature of the neighboring species, as well as the bond-length disorder parameter, or mean-square relative displacement (MSRD). The technique is selective to the absorbing atom, which allows one to probe the local structure around different constituent elements independently.

X-ray absorption near-edge structure (XANES), which involves multiple scattering, additionally allows one to probe the local arrangement of atoms on a scale somewhat beyond the first-nearest neighbors, in particular, it is sensitive to the mutual arrangement of the neighboring atoms in space, i.e. includes bond-angle information. As XANES features are also a consequence of transitions from occupied core states to unoccupied conduction-band states, the spectra also contain information about the density of unoccupied conduction-band states. It should be mentioned that recent advances in theory have made it possible to simulate EXAFS and XANES spectra with good accuracy [49.3].

The Crystalline State

The stable crystal structure of GST is hexagonal [49.4,5]. Thin films, however, crystallize into a different structure. Recent X-ray diffraction (XRD) studies have lead to a conclusion that thin GST layers crystallize into the rock-salt structure with Te atoms oc-

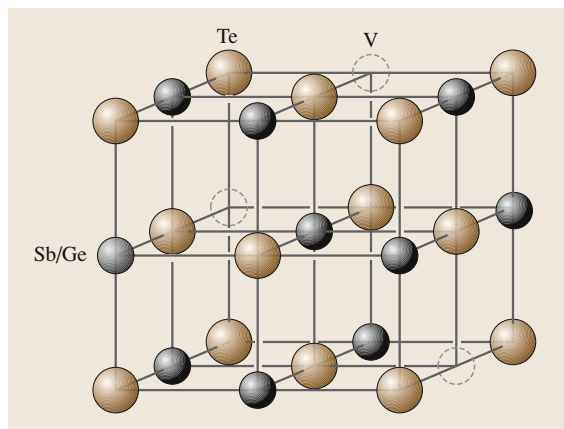


Fig. 49.3 Structure of metastable crystalline Ge–Sb–Te as suggested based on X-ray diffraction measurements

cupying sites on one face-centered cubic (fcc) sublattice with Ge and Sb randomly forming the other fcc sublattice (20% of the sites being vacant) (Fig. 49.3) [49.6, 7]. A lattice parameter of 6.02 Å was reported. The isotropic atomic displacements B_0 , which is a measure of atomic displacements from the ideal crystallographic positions due e.g. to thermal vibrations, found via the fitting process were 1.2 Å² and 3.2 Å² for the Te, and Ge(Sb) species, respectively, which correspond to atomic displacements of 0.1 Å and 0.2 Å for the Te and Ge(Sb) species, respectively.

It was suggested [49.5] that the cubic structure of GST (which is rather isotropic and hence more similar to the amorphous structure than any other crystal structure)

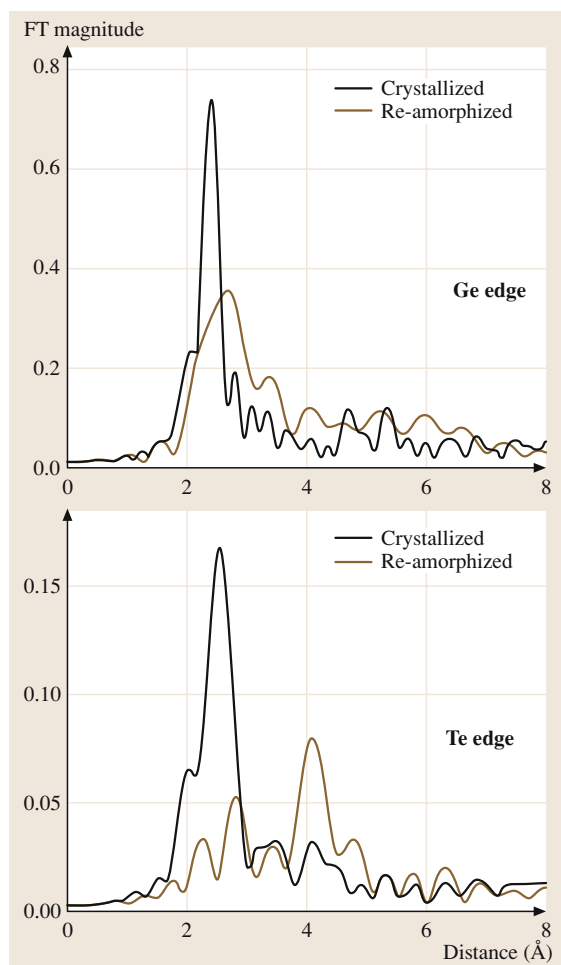


Fig. 49.4 Fourier-transformed Ge and Te K-edge EXAFS spectra of Ge₂Sb₂Te₅ measured for both the crystalline and amorphous states

was the reason for the high-speed switching and stable performance.

The above result gives rise to one more question. Why is B_0 so large (7% of the bond length)? In other words, is the structure really rock salt?

Below we summarize the authors' recent efforts to investigate the local structure of GST using XAFS. Measurements were performed at BL12C at the Photon Factory (Tsukuba, Japan) and BL01B1 at SPring8 (Hyogo-ken, Japan).

The Fourier-transformed (FT) spectra for the Ge and Te edges of GST are shown in Fig. 49.4. It should be noted that the r -space data shown in Fig. 49.4 are not real-space radial distribution function data but the magnitude of the Fourier transforms (FTs) of the k -space EXAFS data. The peak positions in the figure are shifted from the actual interatomic distances toward lower r because of the photoelectron phase shift $\delta(k)$ in the phase factor of the EXAFS oscillations. The spectra measured at the Sb edge did not show any significant variation between the two states and are not shown here.

Details of the data analysis can be found elsewhere [49.8]. The main results for crystalline GST are summarized below. We found two types of bond lengths, namely shorter bonds and longer bonds for both Te–Ge and Te–Sb (for Te–Ge: 2.83 ± 0.01 Å, and 3.2 ± 0.3 Å and for Te–Sb: 2.91 ± 0.01 Å and 3.2 ± 0.3 Å). It should be noted here that the observation of splitting of the bond lengths into two groups is very similar to the case of GeTe. The uncertainties for the longer bonds are rather large. For this reason no definitive conclusions could be drawn about the longer bonds and in what follows we shall exclusively concentrate on the shorter bonds. It should be noticed here that another commercially used material, AIST, also possesses subsets of shorter and longer bonds [49.9].

No Sb–Ge bonds were detected, in agreement with the fact that Sb and Ge do not intermix in the solid phase but we clearly observed a second-nearest-neighbor Te–Te peak at 4.26 Å.

It should also be mentioned that the mean-square relative displacements (MSRD) of the Te–Ge bond length obtained in EXAFS are considerably (lower (0.02 Å²) than) the isotropic atomic displacements of single atoms obtained from XRD (0.04 Å²). This result demonstrates that Ge and Sb atoms do not deviate from the ideal rock-salt positions in a random way but in a strongly correlated manner with respect to the neighboring Te atoms, i.e. the crystalline structure is in fact a *distorted* rock-salt-like structure similar to the case of the ferroelectric GeTe. The off-center location of the Ge atoms

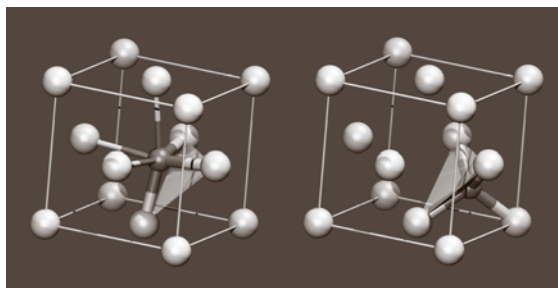


Fig. 49.5 Structural modification of GST upon transformation from the crystalline (*left*) to amorphous (*right*) state (after [49.8])

means that there is a net dipole moment and suggests that GST is a ferroelectric material [49.10].

The Amorphous State

We now turn to the amorphous state. It was found that both Te–Ge and Te–Sb bonds get *shorter* (2.61 Å and 2.85 Å, respectively) and *stronger* upon amorphization, as evidenced by Fig. 49.4. At the same time, the Te second-neighbor peak becomes considerably weaker but does not disappear completely. The MSRD value decreases from 0.02 Å² in the crystalline state to 0.008 Å² in the amorphous state.

Such a behavior is highly unusual for typical three-dimensional covalently bonded solids as, due to the anharmonicity of interatomic potentials, disordering typically results in an increase of the bond lengths and the bond-length disorder. The obtained results reminds one of the case of molecular solids where the presence of intermolecular and intramolecular bonds determines the crystallization–amorphization behavior.

In the current case, a bond-strength hierarchy also exists and the following model of structural rearrangement can be envisaged. Upon melting, the longer Te–Ge (and Te–Sb bonds) are broken and, as a result, the shorter bonds become even shorter and stronger, i.e. the amorphous phase is *locally* more ordered than the crystalline phase. Our Raman scattering experiments provided further grounds for this model, namely, the Raman measurements for both GeTe and GST [49.11] showed that the spectra for the crystalline films are dominated by a peak located at lower wavenumbers, i.e. “mode softening” takes place upon crystallization. This situation can be compared with the case of Se or Te when the interchain interaction is weakened, giving rise to a Raman peak located at higher wave number [49.12]. It should be mentioned that an increased local bond or-

der in the amorphous phase was also observed for selenium [49.13].

To get further insight into the structure of the amorphous phase, we also performed XANES simulations and found that the best agreement with experiment was obtained when Ge was allowed to acquire its preferred tetrahedral surrounding in the amorphous phase [49.8, 11].

This structural transformation is illustrated in Fig. 49.5 where a Ge atom is shown within the fcc sublattice formed by Te atoms. The Ge atoms occupy octahedral and tetrahedral symmetry positions in the crystalline and amorphous states, respectively. The stronger covalent bonds are shown with thicker lines than the weaker bonds (Fig. 49.5 left). An intense laser pulse induces rupture of the weaker bonds and the Ge atom flips into the tetrahedral position (Fig. 49.5 right). An alternative description of the structural transformation upon melting is an umbrella-flip distortion resulting in disordering of the Ge sublattice. Notice, that the three covalent bonds remain intact. This conservation of the system of stronger covalent bonds is crucial: the material is not molten in a conventional sense.

Support for the aforementioned transformation comes from an estimate of the Ge–Te distance from the crystallographic data. Using a lattice parameter of GST obtained by X-ray diffraction [49.6, 7]), the Ge–Te distance—the Ge atoms being in a tetrahedral symmetry position—can be easily calculated to be 2.61 Å, i. e. exactly the value obtained from the EXAFS analysis. This consistency between the results obtained using two different structural techniques is the ultimate proof of the suggested structural modification as well as the generality of the structural modification in GeTe-based alloys.

It is interesting to note that very similar bond lengths for the crystalline and amorphous states were also observed for the binary GeTe [49.11, 14, 15], indicating that it is the GeTe component of the quasibinary

GeTe–Sb₂Te₃ that is mainly responsible for the observed phase transition.

Sb-edge XANES does not exhibit any significant changes upon amorphization (except for the Sb–Te bond shortening [49.8]) implying that the local arrangement of atoms around Sb remains essentially unchanged in accordance with the above model. We believe that the Sb atoms mainly play the role of enhancing overall stability of the metastable crystal structure by participating in the overall electron balance.

Intermediate-Range-Order Changes

The structural change on an *intermediate-range-order* scale can be viewed as following (Fig. 49.6). After rupture of the weaker Ge–Te bonds the Ge atoms flip into the tetrahedral symmetry position forming the GeTe₄ tetrahedra. At the same time, the broken weaker Ge–Te bonds no longer counterbalance the Sb–Te bonds on the opposite site and, as a result, the Sb–Te bonds become structure-determining. The structure relaxes making the Sb–Te bonds shorter (just as in amorphous Se the intrachain bonds get shorter upon amorphization [49.13]). The Sb–Te bond shortening upon amorphization has indeed been observed experimentally [49.8]. This can be viewed as the local phase separation into GeTe and Sb₂Te₃ phases. Finally, the structure relaxation causes a distortion in the Te fcc sublattice. It is important that the amorphous phase possesses a well-defined (single-state) structure without long-range periodicity. We believe that the well-defined local structure of the amorphous state is the reason for the overall stability of the GST-based optical media.

49.1.3 Related Issues

Although it is generally assumed that the role of a laser pulse is simply to heat the material, we believe that electronic excitation creating nonequilibrium charge carriers is crucial for the weakening and subsequent rupture of the subsystem of weaker Ge–Te bonds. Indeed, the presence of longer (weaker) bonds implies that the density of states corresponding to these bonds is lower in energy and hence photogenerated nonequilibrium carriers more readily populate these states, making them more susceptible to thermal vibration-induced dissociation.

As mentioned above, another commercially used material, AIST, has a similar bond hierarchy [49.9], which allows us to suggest that the basic mechanism underlying the phase-change transition in the two groups of materials are essentially the same.

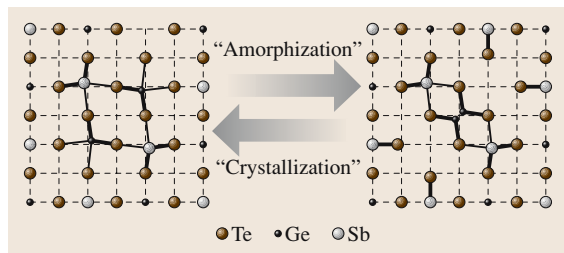


Fig. 49.6 Medium-range-order changes in GST upon crystallization–amorphization transition

It should be stressed that this structural transformation involves a change in the hybridization from p-type bonding in the rock salt to sp^3 hybridization in the amorphous state. This substantial change in the electronic states accounts for the very large change in optical and electrical properties of GST upon the crystallization–amorphization transition.

It should also be noticed here that the unveiled structural modification can also account for why it is the crystalline regions that possess a smaller gap and higher reflectance. Indeed, the crystal structure involves both

shorter and longer bonds. The longer (and hence weaker) bonds are characterized by a smaller energy splitting between the bonding and antibonding states and thus determine the positions of the valence and conduction band edges (and the width of the band gap) in the crystal. In the amorphous phase, on the other hand, the initially weaker bonds break and disappear altogether and the initially stronger bonds get even shorter and stronger. As a result, the splitting between the bonding and antibonding states increases, which produces a larger band gap and lower reflectivity compared to the crystalline phase.

49.2 Super-RENS Discs

Recent progress in multimedia requires continuously higher storage densities and faster data access. A fundamental obstacle here is the so-called diffraction limit, which does not allow one to focus the beam into a spot smaller than $\Phi > \lambda/(2NA)$, where Φ , λ , and NA are the laser spot diameter, light wavelength and numerical aperture (< 1.0) of the lens, respectively.

Various approaches to tackle the problem have been suggested. The most promising among them (within optical data-storage limits) are the following: use of dual-layer disks, use of blue and blue–violet lasers (with a shorter wavelength), use of large numerical apertures, and use of the near-field technique. The latter technique allows one to avoid completely the restrictions on the dot size imposed by the diffraction.

Near-field recording is now being actively developed, following a different routes. One such route is the use of so-called super-resolution near-field structures (super-RENS) [49.16], recently developed at our center. The most important feature of this approach is that the near-field aperture is incorporated into the disk itself, i. e. the near-field recording is realized based solely on thin-film technology. Typical super-RENS structures are shown in Fig. 49.7.

As can be seen from the figure, in addition to the recording layer sandwiched between two capping layers there is a masking layer present. At the initial stages of super-RENS development the masking layer used was Sb (Fig. 49.7a). When an intense laser pulse hits the structure, the Sb layer heats up. When heated, the Sb undergoes the Peierls transition, i. e. a gap in the density of states is formed. As a result, a small aperture opens in it, which serves as a near-field light source [49.16]. The distance between the light source and the recording media is now fixed and determined by the thickness of the capping layer between the masking layer and the GST layer. With optimized parameters for each layer, the aperture in the Sb layer opens for sufficient time to record a bit into the GST layer. When the disk rotates and the exposed Sb spot moves away from the light, the aperture closes. A similar process takes place during the read-out stage.

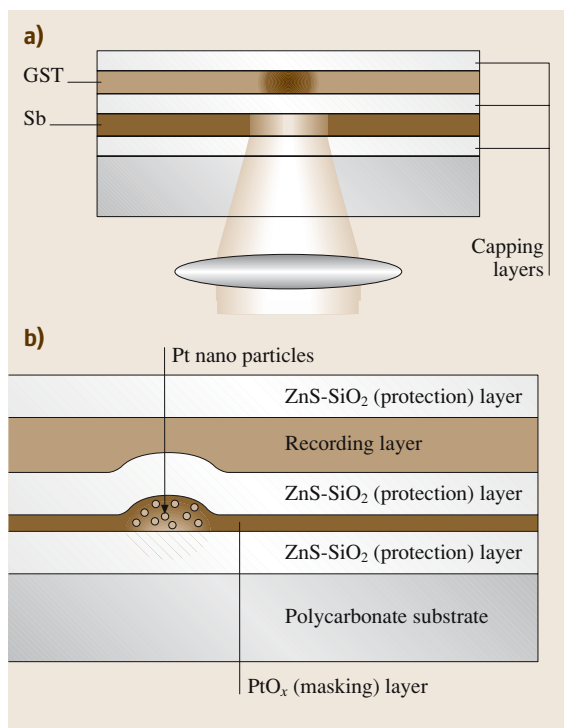


Fig. 49.7 A schematic presentation of (a) transmission- and (b) scattering-type super-RENS discs

An alternative design contains a MeO_x ($\text{Me}=\text{Ag}$, Pt) as the masking layer. Light that is absorbed by the recording layer generates heat and thus causes decomposition of MeO_x with the subsequent formation of Me nanoparticles within an oxygen bubble (Fig. 49.7b) [49.17]. The role of the localized plasmons excited in the nanoparticles has been discussed in detail previously [49.18]. In what follows we very briefly discuss the role of the recording layer.

There are various questions regarding the write-in and read-out mechanisms of super-RENS. In particular, an important question is how can a laser beam with a diameter of $\approx 1\ \mu\text{m}$ read out individual bits with sizes on the order of 100 nm? While there is no definite answer to this question several alternatives have been suggested. Below we describe one possibility, viz., a model based on the ferroelectric catastrophe [49.10].

As discussed earlier, GST in its metastable crystalline form possesses a net dipole moment, i.e. it is a potential ferroelectric. Ferroelectrics are known to possess a very high dielectric constant in a rather narrow temperature range close to the Curie temperature. Intense light acting upon the disk during the read-out heats the rotating disk. Because of the rotation, the temperature profile is asymmetric. There is only a rather

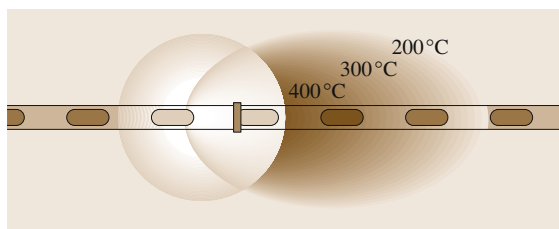


Fig. 49.8 The ferroelectric catastrophe model of super-RENS read-out

narrow region where the temperature of the GST corresponds to the Curie temperature. It is only this region that possesses a high dielectric constant and a very high reflectivity, which enables one to read out marks whose size is considerably smaller than the laser spot [49.10].

Interestingly, the other phase-change material, AIST, is also one of the best materials for use in super-RENS discs. Since both kinds of materials possess longer and shorter bonds, and hence a net dipole moment, the suggested ferroelectric-catastrophe model may operate in both cases. Experiments to further investigate the ferroelectric properties of GST and AIST and their relevance to the super-RENS read-out mechanism are currently underway.

49.3 In Lieu of Conclusion

While it took almost 40 years to elucidate the origin of the structural change behind conventional phase-change recording, we hope that progress related to understanding of the super-RENS mechanism will be achieved on a much shorter timescale. To complete this chapter, we would like to note that presently

marks as small as 80 nm can be successfully recorded with a carrier-to-noise ratio (CNR) over 40 dB. The progress obtained in the field so far makes us confident that optical recording at a density higher than 100 GB/disc will become widely available in the very near future.

References

- 49.1 R. Ovshinsky: Phys. Rev. Lett. **21**, 1450 (1968)
- 49.2 T. Ohta, S. R. Ovshinsky: *Photo-Induced Metastability in Amorphous Semiconductors*, ed. by V. Kolobov A. (Wiley-VCH, Weinheim 2003) p. 310
- 49.3 A. L. Ankudinov, B. Ravel, J. J. Rehr: Phys. Rev. B **58**, 7565 (1998)
- 49.4 I. I. Petrov, R. M. Imamov, Z. G. Pinsker: Sov. Phys. Cryst. **13**, 339 (1968)
- 49.5 N. Yamada, T. Matsunaga: *International Symposium on Optical Memories. Technical Digest* (Jeju, Korea 2004) p. 70
- 49.6 T. Nonaka, G. Ohbayashi, Y. Toriumi, Y. Mori, H. Hashimoto: Thin Solid Films **370**, 258 (2000)
- 49.7 N. Yamada, T. Matsunaga: J. Appl. Phys **88**, 7020 (2000)
- 49.8 A. V. Kolobov, P. Fons, A. I. Frenkel, A. L. Ankudinov, J. Tominaga, T. Uruga: Nature Mater. **3**, 703 (2004)
- 49.9 T. Matsunaga, Y. Umetani, N. Yamada: Phys. Rev. B **64**, 184 116 (2001)
- 49.10 J. Tominaga, T. Shima, M. Kuwahara, T. Fukaya, A. V. Kolobov: Nanotechnology **15**, 411 (2004)
- 49.11 A. V. Kolobov, P. Fons, J. Tominaga, A. L. Ankudinov, S. Yannopoulos: J. Phys.: Condens. Matter **16**, S5103 (2004)

- 49.12 M.H. Brodsky: *Light Scattering in Solids*, ed. by M. Cardona (Springer, Berlin Heidelberg New York 1983) p.205
- 49.13 D.E. Sayers: Structural Studies of Disordered Systems using EXAFS. In: *Proc. 7th Int. Conf. Amorph. and Liquid Semicond.*, ed. by W.E. Spear (University of Edinburgh, Edinburgh 1977) p.61
- 49.14 A.V. Kolobov, P. Fons, J. Tominaga: Appl. Phys. Lett. **82**, 382 (2003)
- 49.15 Y. Maeda, M. Wakagi: Jpn. J. Appl. Phys. **30**, 101 (1991)
- 49.16 J. Tominaga, T. Nakano, N. Atoda: Appl. Phys. Lett. **73**, 2078 (1998)
- 49.17 T. Kikukawa, T. Nakano, T. Shima, J. Tominaga: Appl. Phys. Lett. **81**, 4697 (2002)
- 49.18 J. Tominaga, D. P. Tsai (Eds.): *Optical Nanotechnology: The Manipulation of Surface and Local Plasmons*, ed. by J. Tominaga, D. P. Tsai (Springer, Berlin Heidelberg New York 2003) p.212

50. Carbon Nanotubes and Bucky Materials

The chapter details the underlying phenomena that underpin electronic applications that have followed from the discoveries of C_{60} and carbon nanotubes. The reduced dimensionality of these self-organised structures, high electron mobility, weak electromigration, and the plethora of quantum electronic effects exhibited by these structures suggest they are serious candidates for molecular electronics. The detail of the surface chemistry and conditions of synthesis assume greater importance than for conventional electronic materials since all atoms are on the exterior of these structures, as is outlined with references to the wider literature. Essential electronic structure information is given with reference to the transport measurements that have contributed greatly to the evolution of

50.1 Carbon Nanotubes	1147
50.1.1 General	1147
50.1.2 Geometry	1149
50.1.3 Synthesis and Chemistry	1149
50.1.4 Electronic Structure and Transport	1150
50.1.5 Nanoelectronic Devices	1151
50.1.6 Other Electronic Applications	1152
50.2 Bucky Materials	1153
References	1153

the field with emphasis on the Coulomb blockade and ballistic transport phenomena. The major electronic applications are then outlined, giving the state-of-the-art figures of merit for performance and comments on prospects for realisation.

The discovery of C_{60} (Fig. 50.1) in 1986, and subsequently higher fullerenes, followed by that of carbon nanotubes in 1991 accelerated the interest in carbon-based molecules and compounds as electronically or optically active materials that had begun in the early

1970s. The emergence of the interdisciplinary field of nanotechnology has beneficially attracted attention and methodologies from other fields of science and engineering that have considerably enhanced the physical electronics research effort [50.1, 2].

50.1 Carbon Nanotubes

50.1.1 General

The electronic and structural description of carbon nanotubes and other sp^2 -bonded carbon nanostructures with curved surfaces (bucky materials, the term ‘bucky’ is derived from buckminsterfullerene: the name given to C_{60} soon after discovery) will inevitably contain references to that of the familiar two-dimensional (2-D) planar counterpart, graphite. The strength of the in-plane covalent bond of the hexagonal graphitic network produces a short C–C distance of 0.142 nm and high mechanical stiffness.

Carbon nanotube is a generic term that includes both single- and multi-walled structures. Single-wall carbon

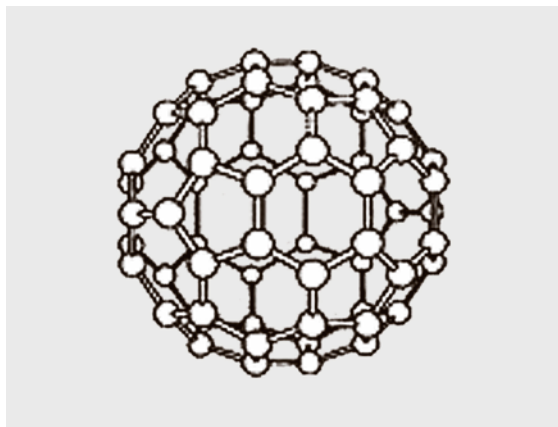


Fig. 50.1 The C_{60} molecule, diameter 0.7 nm

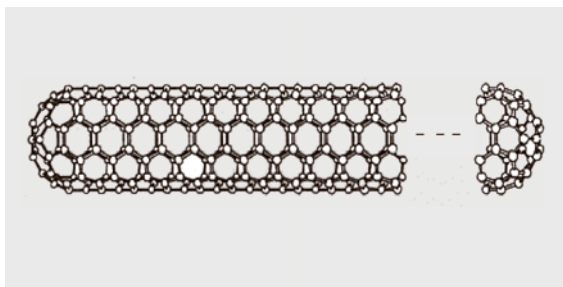


Fig. 50.2 Single-wall carbon nanotube, length 1–10 μm , diameter typically 1.4 nm to > 10 nm

nanotubes (SWNTs) are seamless cylindrical graphitic macromolecules of nanoscale diameter and micrometer length, capped by hemispherical ends. The closure of the cylinder is a result of pentagonal inclusions in the hexagonal carbon network of the nanotube walls, Fig. 50.2. For SWNTs, every carbon atom is a surface atom. The smallest SWNT diameter reported to date, 4 Å, corresponds to the predicted lower limit for stable SWNT formation from consideration of the stress energy built into the cylindrical structure [50.3]. SWNTs tend to form closely packed bundles, or *ropes*, of several tens of individual single-wall nanotubes.

Multi-wall nanotubes (MWNTs) comprise several to tens of concentric cylinders of these graphitic shells with a layer spacing of 3.4 Å (the inter-planer spacing of graphite). SWNT diameters are typically ≈ 1 nm and a MWNT diameter can be in the range 2–100 nm with typical values of 10–20 nm. Nanotube lengths are typically 1–10 μm but can be ≈ 1 mm, giving rise to astonishing aspect ratios (length/diameter) of > 1000 . MWNTs of similar dimensions can also be formed from a single graphite sheet rolled into a scroll structure [50.4]. Double-wall carbon nanotubes are a recent addition to the set of carbon nanostructures.

A consequence of the high C–C bond strength is that the in-plane Young's modulus of a carbon nanotube is $\cong 1$ TPa, making the carbon nanotube one of the stiffest known materials, while the tensile strength is $\cong 150$ GPa (nanotubes have 600 times the strength/weight ratio of steel). Nanotubes deform elastically by buckling, have very low defect density over hundreds to thousands of interatomic spacings, and the strain energy built into the cylindrical structure tends to promote self-repair. The stiff nanotube structure means that the optical-phonon population is minimal even at room temperature; the result can be ballistic electron transport over micrometer length scales (several orders of magnitude greater than that in conventional semiconductor nanostructures).

Another consequence of the short C–C distance is spatially extensive overlapping hybridised atomic orbitals – a π -electron system – with highly mobile delocalised electrons on the interior and exterior of the structure. One carbon atom contributes one unpaired π -electron to the nanotube. The promise of very high carrier mobility in the π -electron system is one motivation for carbon-based electronics. Another is the range of electronic attributes that are displayed by the carbon allotropes, e.g. room-temperature resistivity can range from $10^{-8} \Omega\text{m}$ for SWNTs to $> 10^{18} \Omega\text{m}$ for crystalline C_{60} .

The π -electron system and stiff structure produce electrical conductivity, very high current-carrying capacity with weak electromigration. The best figures of merit for nanotube systems that have been measured since discovery are: electrical conductivity, $10^8 \Omega^{-1}\text{m}^{-1}$ (comparable with that of copper), carrier mobility: $10^4 \text{cm}^2/\text{Vs}$ (comparable with that of high-quality GaAs), thermal conductivity 10^4W/mK (comparable with that of diamond), and sustainable current density of 10^{13}A/m^2 (> 1000 times greater than that of a metal nanowire).

Nanotube circumferential crystal momentum vectors are quantised due to the periodic boundary conditions imposed by the finite nanoscale diameter whereas those directed along the micrometer-scale axis show continuous variation; thus, one-dimensional (1-D) electronic transport is intrinsic for the SWNT. Electronic systems of reduced dimensionality – dots, wires and sheets of charge – created by top-down silicon processing have assumed importance in conventional microelectronics. Carbon nanostructures offer the possibility of self-organised molecular-level low-dimensional electronic systems, e.g.: zero-dimensional (0-D) (C_{60} and short SWNTs), 1-D (SWNTs), 2-D (large-diameter MWNT outer shell), and three-dimensional (3-D) (coupled-layer MWNTs). Electronic modification of carbon nanotubes by insertion of C_{60} and other fullerenes into the central capillary to form a mixed-dimensionality hybrid structure called the *peapod* is a science in its infancy but promises a rich future for research and applications [50.5].

Carbon nanotubes are chemically stable in ambient conditions and structurally stable in vacuum for temperatures far greater than 1000°C . Viewed in total, the above properties potentially make carbon nanotubes excellent electronic materials. For some nanoelectronic applications the major drawback is perhaps insufficient control of the synthesis process – presently the focus of much worldwide research effort – and contacting of in-

dividual nanotubes is poorly understood. Nevertheless, the range of proposed electronic applications is vast. These considerations place carbon nanotube electronics among a number of competing technologies poised to complement or replace silicon-based complementary metal–oxide–semiconductor (CMOS) technology in the < 100 nm-feature-size domain.

50.1.2 Geometry

The unique feature of the electronic structure of carbon nanotubes is that it can range from the metallic to the semiconductor depending on the details of the microstructure. Essentially, the way the hexagonal network connects to itself to form a cylinder determines the electronic structure: the chiral vector usually used to describe the wrapping of the network is

$$\mathbf{C} = n\mathbf{a}_1 + m\mathbf{a}_2 \equiv (n, m),$$

where \mathbf{a}_1 and \mathbf{a}_2 are the unit vectors of the hexagonal network, and n and m are integers. The resulting nanotube can then be described in the form (n, m) . Electronic band-structure calculations show that an (n, m) nanotube is metallic at room temperature if $2n + m$ is a multiple of 3, otherwise it is a semiconductor with a band gap of $E_g = 0.9 d \text{ eV}$, where d is the nanotube diameter in nanometers, i. e. typically $E_g \approx 0.5 \text{ eV}$. This remarkable property signals the possibility of band-gap engineering by control of the microstructure. The wrapping angle, or chiral angle, is given by the angle between \mathbf{a}_1 and \mathbf{C} . From geometric and symmetry considerations, $0 < |m| < n$, $0^\circ = \theta = 30^\circ$, $\cos(\theta) = 2n + m / [2\sqrt{n^2 + m^2 + nm}]$, and nanotube diameter $d = 0.078\sqrt{n^2 + m^2 + nm}$ nanometers. The general case (n, m) is referred to as the chiral nanotube; there are two special cases: (a) the zigzag nanotube ($m = 0$, $\theta = 0^\circ$), and (b) the armchair nanotube ($n = m$, $\theta = 30^\circ$, all metallic) [50.2].

Some degree of control of diameter and length is possible at present but very limited control of chirality, and therefore electronic structure, is achievable now. Assuming a random distribution of nanotube diameters in the reaction products, the relative populations of metallic to semiconducting electronic structure is 1:3. Single-symmetry (n, m) synthesis or post-synthesis sorting according to electronic structure are key fields for research effort. The task is proving rather difficult but there appear to be no fundamental barriers to progress.

A stable defect in the hexagonal network is the pentagon–heptagon pair (Stone–Wales or 5–7 defect). This defect will cause a sharp bend in an otherwise well-graphitised SWNT. Controlled introduction of defects

allows the possibility of constructing Y- and T-junctions and other complex geometries from nanotubes, including rings and coils. Moreover, a 5–7 defect can connect a metallic to a semiconductor nanotube giving an Angstrom-scale heterojunction and hence a device density 10^4 times greater than present-day microelectronics [50.6].

50.1.3 Synthesis and Chemistry

The quality of carbon nanotubes in terms of crystallinity and impurity content depends on the conditions of synthesis by arc discharge using graphitic electrodes, laser vaporisation of graphite, or chemical vapour deposition (CVD) using hydrocarbon gas and metal nanoparticle catalysts. Each method produces different nanotube samples and sample-to-sample variation from the same source. An extensive review of synthesis methods was published by Rakov [50.7]. The key factor for crystallinity, and therefore electronic quality, is the formation temperature: $> 2000^\circ\text{C}$ is required for complete graphitisation. Little is known about the preferences the various growth processes have for certain nanotube symmetries over others.

Reaction products can contain significant quantities of residual metal catalyst particles and non-tubular forms of carbon (typically 30 wt % of the raw material). Purification processes ranging from the simple to complex are commonly used: oxidation and acid washing [50.8], solvent treatment followed by ultrafiltration [50.9], or flocculation using aqueous surfactants [50.10]. Controlled growth of MWNTs perpendicular to a substrate has been widely explored due to its technological importance [50.11].

Surface functionalisation of SWNTs is required for several applications as a means of modifying charge exchange with the ambient. Oxidation of nanotubes with the acid mixture $\text{H}_2\text{SO}_4\text{--HNO}_3$ leads to high concentrations of carboxylic, carbonyl, and hydroxyl groups on the surface and removal of the tips to expose the interior surface. Carboxyl ($-\text{COOH}$) groups are then readily derivatised by a variety of reactions. Covalent functionalisation, however, necessarily disrupts the rigid structure and π -electron system with consequent degradation of mechanical and electrical properties but provides the best stability and accessibility [50.12]. Non-covalent routes to nanotube functionalisation offer ease of synthesis and minimum disruption of the tubular structure [50.13].

Exposure to ambient humidity, oxygen and other gases profoundly affects the measured electronic prop-

erties of nanotubes through charge exchange and the quality of metallic electrical contacts made to nanotubes. The electronic properties of carbon nanotube systems are extremely sensitive to the presence of molecular oxygen due to the formation of the charge-transfer complex $C_p^{+\delta}-O_2^{-\delta}$, i. e. oxygen-exposed nanotubes are p-type [50.14]. Nanotubes exhibit ultrahigh sensitivity at room temperature to O_2 , N_2O , NH_3 but not H_2 [50.15]. Nanotubes must be elevated to temperatures $> 700^\circ C$ in ultrahigh vacuum to completely remove absorbents.

50.1.4 Electronic Structure and Transport

The electronic structure of SWNTs is predicted by tight-binding calculations extended from the 2-D approach for graphite to the cylindrical 1-D nanotube system [50.2]. In the case of nanotubes circumferential crystal momentum vectors are quantised due to the periodic boundary conditions imposed by a finite nanoscale diameter. This simple approach has been proven, mainly by low-temperature scanning tunnelling spectroscopy, to be a good description of the electronic structure of carbon nanotubes [50.16]. The curvature of the nanotubes is ignored in this model but can introduce some important modifications [50.17]. However, the original model is, nevertheless, adequate for the classification of nanotubes into metallic and semiconducting types according to the simple rule outlined in Sect. 50.1.1. In common with other 1-D conductors with half-filled bands the expectation was that carbon nanotubes will undergo a Peierls distortion, i. e. and a gap will open in the band structure at the Fermi energy due to atomic displacement modifying the unit cell. However, it has been shown theoretically that the effect very quickly diminishes as a function of nanotube radius [50.2].

Electrical transport measurements on individual SWNTs, SWNT ropes, and MWNTs have produced diverse results in the decade since discovery and are still the focus of intense research effort. The picture is complicated by several factors:

1. the current pathway through MWNTs is greatly complicated by interlayer coupling,
2. two-terminal resistances in the range $100\ \Omega$ – $100\ M\Omega$ were first reported [50.18],
3. the processing, contacting, and manipulation methods used undoubtedly introduced non-intrinsic resistances and potential barriers.

Nevertheless, the best transport parameters that have emerged since discovery, quoted in Sect. 50.1, are highly

desirable from many device points of view. The quality of lithographically defined contacts is critical for electrical transport measurements. There are three possible outcomes of the contacting process:

1. the contacts are of good quality; then the total resistance will be the diffusive or quantum resistance for that structure ($6.5\ k\Omega$ for a ballistic SWNT) [50.19];
2. no contact is made and the nanotube is isolated by tunnel barriers; in this case the nanotube will act as a Coulomb island and blockade phenomena will be observed at low temperature [50.20],
3. the contacts are of an intermediate type that allows the nanotube to act as a cavity and Fabry–Perot interference phenomena are observed [50.21].

The exact physics of contacting to carbon nanotubes is poorly understood; the exact atomic arrangement at the interface is likely to play a role as are the ambient gases.

Successful transport measurements have yielded a plethora of quantum-electronic and low-dimensional transport phenomena. *Tans* et al. [50.19] demonstrated true quantum wire behaviour; the remarkable feature of this work was to identification of an electron phase-coherence length on the micrometer scale at 4 K for nanotubes: several orders of magnitude greater than that in conventional semiconductor nanostructures. With an extraordinarily skilful experiment *Bachtold* et al. [50.22] demonstrated the Aharonov–Bohm effect by circulating circumferential current in a contacted MWNT with the axis parallel to a magnetic field. The periodicity of the magneto-conductance variation with applied field corresponded to an interferometer diameter equal to that of the outer MWNT shell. This finding gave weight to an earlier conclusion from quantum conductance measurements of *Frank* et al. [50.23], namely that current is only carried in the outer shell for MWNTs. This was posed as a general statement but was contradicted by the controlled layer-by-layer conductivity analysis of *Collins* et al. [50.24] in which the current-carrying contribution of each MWNT layer was determined by vaporisation of successive MWNT layers. This work concluded that at least eight layers contribute to the current-carrying capacity of a MWNT. The nature of MWNT inter-layer coupling is a point of discussion at present.

The landmark paper of *Frank* et al. identified ballistic conductance in MWNT systems at room temperature by controlled lowering of a MWNT bundle into a liquid mercury contact [50.23]. This was a remarkable result since the implication is that the electron mean free path is on the micrometer scale at room temperature. A ballis-

tic conductor has a conductance given by the Landauer formula:

$$G = 2 \left(\frac{e^2}{h} \right) \cdot \sum_i^N T_i,$$

where N is the number of conducting channels, T is the transmission factor (experimental result: $T = 1$ for most nanotube systems), and $G_0 = 2e^2/h$ is the quantum conductance (the factor of 2 is due to spin degeneracy). Frank et al. measured a conductance of $1G_0 >$ per MWNT, implying that current is carried in the outer shell only and there is a *missing* $1G_0$ since a metallic carbon nanotube has two crossing 1-D energy bands at the Fermi level, which yields a conductance of $2G_0$ for a single nanotube shell. However, more recent results have shown the expected $2G_0$ result [50.21]. Therefore, the controversy over the experiment by Frank et al. remains unsolved.

SWNTs were expected to exhibit the predicted electron–electron interaction effects in 1-D systems, the so called Luttinger liquid (LL) state. In a strictly 1-D system the independent electron approximation breaks down and the normal 3-D Fermi gas picture is replaced by the LL. In the LL description electron–electron interactions dominate and the electrons are in a highly correlated state. The LL state requires a high degree of order since disorder destroys the correlated ground state. The main features of the LL are suppression of the density of states at the Fermi energy according to a power law and separation of spin and charge excitations. In transport measurements, the LL state is manifested as a tunnelling conductance from a normal electrode according to $G(T) \approx T^\alpha$, for $eV \ll k_B T$, and at large voltages $G(V) = dI/dV \approx V^\alpha$ for $eV \gg k_B T$, where the parameter α is determined by the strength of the interaction between the electrons. Such power laws have been observed in contacted SWNTs, although stronger evidence, such as a clear demonstration of spin–charge separation, is necessary to identify the LL state unambiguously [50.25].

Coulomb blockade phenomena have been observed in low-temperature transport measurements on carbon nanotube systems often by unintentional incorporation of tunnel junctions between the metal electrode and the nanotube. The effect occurs when a mesoscopic *island* with capacitance C (self-capacitance of the nanotube plus the barrier capacitance at the nanotube–electrode junction) is isolated by tunnel barriers from two electrodes. If the charging energy for a single electron on the island, e^2/C , is larger than the thermal energy $k_B T$, then

the system will not conduct due to electron–electron repulsion. Coulomb blockade can be modulated by a gate electrode from full blocking to a conducting state limited only by the tunnel barriers. Such a system forms a single-electron transistor (SET). The capacitance of a nanotube is $\approx 10^{-17}$ F, thus SET operation conditions are fulfilled at temperatures below 10 K when the entire nanotube defines the Coulomb island. To minimise the capacitance and therefore elevate the operating temperature inter-tube islands can be defined by crossing nanotubes or intra-tube islands bending with a scanning probe microscope (kinks and bends in SWNTs can create tunnel barriers). The latter approach has been used to create a SET that operates close to room temperature [50.26].

Superconductivity in nanotube systems was first observed as the proximity effect with SWNTs or SWNT ropes connecting two superconducting electrodes [50.27]. The first report signalling intrinsic superconductivity in SWNT ropes was a recent landmark publication [50.28]. This will clearly spur further experimental and theoretical effort.

50.1.5 Nanoelectronic Devices

Semiconducting nanotubes are especially important for nanoelectronic device applications. Semiconducting behaviour has been observed in individual SWNTs; in SWNT ropes and MWNTs mixed metallic and semiconducting current pathways greatly complicate the I – V characteristic. Field-effect transistors (FETs) with individual SWNT channels have been demonstrated and found to have higher transconductance than can be achieved with state-of-the-art metal–oxide–semiconductor field-effect transistors (MOSFETs) [50.29]. The conducting state is attained with negative gate voltages, implying that the carbon nanotube forms a normally off p-type conduction channel (probably due to unintentional doping by exposure to oxygen). Logic gates made from nanotube FETs have been recently demonstrated [50.30]. The nature of the metal electrode–SWNT Schottky barrier is still not fully understood.

The prospects for highly integrated circuits consisting of nanotube elements must be considered remote at this time for several reasons (some previously mentioned):

1. current synthesis techniques produce mixtures of metallic or semiconductor nanotubes and these tend to form ropes or bundles;

2. the precise nature of the metal electrode–nanotube contact is poorly understood; and
3. scanning probe manipulation techniques used to fabricate prototype individual devices are not scalable.

50.1.6 Other Electronic Applications

Although nanoelectronics has been the driving force for carbon nanotube research, many other electronic applications are being explored; this section outlines some of them.

Interconnects

Ballistic transport on the micrometer scale at room temperature and high current-carrying capacity suggests carbon nanotubes are good candidates for high-bandwidth dissipation-less interconnect for nanoscale circuit elements. Since chemical processing tends to degrade nanotube electrical properties, methods of directed growth are being explored as a means of *wiring* integrated circuits, e.g. guiding through via holes [50.31], by electric field [50.32], or surface modification [50.33].

Field-Emission Displays

The nanoscale diameter, aspect ratio of ≈ 1000 , and high conductivity make carbon nanotubes ideal candidates for field emitters [50.34]. Field-emission currents from single nanotubes and aligned or randomly oriented nanotube thin films have been extensively studied. The promise is of low-threshold-field electron emission with a current density sufficient to drive a phosphor screen for display purposes. SWNTs and MWNTs have proved to be remarkably good field emitters with threshold emission fields in the range $1\text{--}10\text{ V}/\mu\text{m}$ and capable of carrying very high current density. The exact emission mechanism, the role played by surface absorbents, and the cause of emission current saturation are under debate. Nevertheless, major manufacturers have recently produced prototype field-emission displays with carbon nanotube sources; with further development, this is likely to be the first major application of carbon nanotubes to reach the marketplace [50.35].

Electron-Beam Lithography

The present minimum feature size for silicon microelectronics is 130 nm, which is achieved using extreme-UV optical lithography. Alternative technologies are being explored for the future production of $< 100\text{ nm}$ devices. Conventional electron-beam lithography can achieve a 5-nm line width but is limited by writing time because a single beam is used to write the entire pattern. Thus

arrays of electron guns operating in parallel are being considered as a route to reduction of the writing time while maintaining the high resolution of electron-beam lithography. Carbon nanotubes are currently the most promising candidates for use as the emission source [50.36].

Electro-Optic Materials

Polymer–nanotube composite materials have been studied from the point of view of applications in electro-optics. Loading the layers of organic light-emitting diodes (LEDs) with low concentrations of nanotubes effectively increased the lifetime of the devices by preventing the build up of local hot spots thorough the high thermal conductivity that can be achieved in a nanotube percolation network [50.37]. Nanotube-induced local ordering of the matrix polymer suggests efficiency improvements may also be possible [50.38]. Using nanotube mats as an electrode in solar-cell applications apparently gave no improvement in device performance; however, these early measurement were performed with defective MWNTs so there may be a case for further investigation using the high-quality SWNTs available today.

Nanofabricators

Carbon-nanotube atomic force microscopy tips for fabricating oxide nanostructures in Si and Ti by anodisation is a rapidly expanding field aimed at Tbit/cm^2 data storage [50.39]. The technique utilises ambient moisture in the oxidation process with the tip biased negative relative to the surface. Line widths of $5\text{--}10\text{ nm}$ suitable for antidotes and tunnel junctions can be achieved.

Actuators

Electromechanical actuators based on SWNT sheets have been shown to generate higher stresses than natural muscle when operating in physiological conditions and higher strains than high-modulus ferroelectrics [50.40]. The actuation mechanism is a geometrical expansion of the carbon–carbon covalent bond caused by electrochemical double-layer charging. Work densities per cycle substantially higher than any previously known technology are predicted for SWNT sheets with mechanical properties close to those of individual SWNTs. This can be achieved by inter-tube binding and alignment optimisation.

Sensors

The conductivity of nanotube systems is highly sensitive to gaseous ambients, which affect the sign and

amount of injected charge. The dimensions of a nanotube sensing element is such that very low quantities of analyte species will produce a measurable response. Nanotube gas sensors certainly have prospects to challenge conventional gas sensors for certain uses [50.15].

50.2 Bucky Materials

Over the last decade the field of bucky materials has been dominated by carbon nanotubes, as the number of publications and patent submissions testify. After intense research activity since their discovery, the most noteworthy of C_{60} -based electronic materials are superconducting heterofullerides of composition K_2MC_{60} , where $M=Fe, Ni, Cu, Ag, Co \dots$, with $T_c \approx 15$ K. However the future for encapsulation of species in the C_{60} , and higher fullerene, cage ap-

In addition to transduction, bio-sensors require a bio-receptor (e.g. enzyme or cell) immobilisation matrix. Carbon-nanotube-based bio-sensors meet both requirements and have been found to promote homogeneous electron-transfer reactions.

pears promising, in particular as the basis of devices for quantum information processing. Other non-carbon-nanotube structures are now under intense investigation; these include BN, W, MoS_2 , heterogeneous nanotubes and the ternary compound $Mo_6C_{9-x}H_x$ (C =chalcogen, H =halogen, $3 < x < 6$). The science of these structures is still in its infancy but, given the rich experience of over a decade of carbon nanotube research, there is reason to be optimistic of exciting new science and technology.

References

- 50.1 M. S. Dresselhaus, G. Dresselhaus, Ph. Avouris (eds): *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*, Topics Appl. Phys., Vol. 80 (Springer, Berlin Heidelberg New York 2000)
- 50.2 R. Saito, G. Dresselhaus, M. S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College Press, London 1998)
- 50.3 N. Wang, Z. K. Tang, G. D. Li, J. S. Chen: *Nature* **408**, 50 (2000)
- 50.4 L. M. Viculus, J. J. Mack, R. B. Kaner: *Science* **299**, 1361 (2003)
- 50.5 X. Liu, T. Pichler, M. Knupfer, M. S. Golden, J. Fink, H. Kataura, Y. Achiba, K. Hirahara, S. Iijima: *Phys. Rev. B* **65**, 45419 (2002)
- 50.6 L. Chico, V. H. Crespi, L. X. Benedict, S. G. Louie, M. L. Cohen: *Phys. Rev. Lett.* **76**, 971 (1996)
- 50.7 E. G. Rakov: *Russ. Chem. Rev.* **69**, 25 (2000)
- 50.8 A. C. Dillon, T. Genett, K. M. Jones, J. L. Alleman, P. A. Parilla, M. J. Heben: *Adv. Mater.* **11**, 1354 (1999)
- 50.9 K. Tohji, H. Takahashi, Y. Shinoda, N. Shimizu, B. Jeyadevan, I. Matuoka, Y. Sato, A. Kasuya, S. Ito, Y. Nishina: *J. Phys. Chem. B* **101**, 1974 (1997)
- 50.10 J.-M. Bonard, T. Stora, J.-P. Salvetat, F. Maier, T. Stoeckli, C. Duschul, L. Forro, W. A. de Heer, A. Chatelain: *Adv. Mater.* **9**, 827 (1997)
- 50.11 A. Huczko: *Appl. Phys. A* **74**, 617 (2002)
- 50.12 S. E. Baker, W. Cai, T. L. Lasseter, K. P. Weidkamp, R. J. Hamers: *Nano. Lett.* **2**, 1413 (2002)
- 50.13 R. J. Chen, Y. Zhang, D. Wang, H. Dai: *J. Am. Chem. Soc.* **123**, 3838 (2001)
- 50.14 G. U. Sumanasekera, C. K. W. Adu, S. Fang, P. C. Ek-lund: *Phys. Rev. Lett.* **85**, 1096 (2000)
- 50.15 J. Kong, N. R. Franklin, C. Zhou, M. G. Chapline, S. Peng, K. Cho, H. Dai: *Science* **287**, 5453 (2000)
- 50.16 J. W. G. Wildöer, L. C. Venema, A. G. Rinzler, R. E. Smalley, C. Dekker: *Nature* **391**, 59 (1998)
- 50.17 M. Ouyang, J. L. Huang, C. L. Cheung, C. M. Lieber: *Science* **292**, 702 (2001)
- 50.18 T. W. Ebbesen, H. J. Lezec, H. Hiura, J. W. Bennett, H. F. Ghaemi, T. Thio: *Nature* **382**, 54 (1996)
- 50.19 S. J. Tans, M. H. Devoret, H. Dai, A. Hess, R. E. Smalley, L. G. Geerlings, C. Dekker: *Nature* **386**, 474 (1997)
- 50.20 M. Bockrath, D. H. Cobden, P. L. McEuen, N. G. Chopra, A. Zettl, A. Thess, R. E. Smalley: *Science* **275**, 1922 (1997)
- 50.21 W. Liang, M. Bockrath, D. Bozovic, J. H. Hafner, M. Tinkham, H. Park: *Nature* **411**, 665 (2001)
- 50.22 A. Bachtold, C. Strunk, J.-P. Salvetat, J.-M. Bonard, L. Farro, T. Nussbaumer, C. Schonenberger: *Nature* **397**, 673 (1999)
- 50.23 S. Frank, P. Poncharal, Z. L. Wang, W. A. de Heer: *Science* **280**, 1744 (1998)
- 50.24 P. C. Collins, M. S. Arnold, P. Avouris: *Science* **292**, 1331 (2001)
- 50.25 M. Bockrath, D. H. Cobden, L. Jia, A. L. Rinzler, R. E. Smalley, L. Balents, P. L. McEuen: *Nature* **397**, 598 (1999)
- 50.26 H. W. Postma, T. Teepen, Z. Yao, M. Grifoni, C. Dekker: *Science* **293**, 76 (2001)
- 50.27 A. F. Morpurgo, J. Kong, C. M. Marcus, H. Dai: *Science* **286**, 263 (1999)
- 50.28 M. Kociak, A. Yu. Kasumov, S. Gueron, B. Reulet, L. Vaccarrini, I. I. Khodos, Yu. B. Gorbakov,

- V. T. Volkov, H. Bouchiat: *Phys. Rev. Lett.* **86**, 2416 (2001)
- 50.29 S. J. Wind, J. Appenzaller, R. Martel, V. Derycke, Ph. Avouris: *Appl. Phys. Lett.* **80**, 3817 (2002)
- 50.30 A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker: *Science* **294**, 1317 (2001)
- 50.31 F. Kreupl, A. P. Graham, G. S. Duesberg, W. Steinhögl, M. Liebau, E. Unger, W. Hönlein: *Microelectron. Eng.* **64**, 399 (2002)
- 50.32 Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, H. Dai: *Appl. Phys. Lett.* **79**, 3155 (2001)
- 50.33 J. Liu, M. J. Cavasant, M. Cox, D. A. Walters, P. Boul, L. Wei, A. J. Rimberg, K. A. Smith, D. T. Colbert, R. E. Smalley: *Chem. Phys. Lett.* **303**, 125 (1999)
- 50.34 J.-M. Bonard, M. Croci, C. Klinke, R. Kurt, O. Noury, N. Weiss: *Carbon* **40**, 1715 (2002)
- 50.35 N. S. Lee, D. S. Chung, I. T. Han, J. H. Kang, Y. S. Choi, H. Y. Kim, S. H. Park, Y. W. Jin, W. K. Yi, M. J. Yun: *Diamond Rel. Mater.* **10**, 265 (2001)
- 50.36 W. I. Milne, K. B. K. Teo, M. Chhowalla, G. A. J. Amaratunga, J. Yuan, J. Robertson, P. Legagneux, K. Pirio, K. Bouzehouane, D. Pribat, W. Bruenger, C. Trautmann: *Curr. Appl. Phys.* **1**, 317 (2001)
- 50.37 B. McCarthy, J. N. Coleman, R. Czerw, A. B. Dalton, M. in het Panhuis, A. Maiti, A. Drury, P. Bernier, J. B. Nagy: *J. Phys. Chem. B* **106**, 2210 (2002)
- 50.38 H. Ago, K. Petritsch, M. S. P. Shaffer, A. H. Windle, R. H. Friend: *Adv. Mater.* **11**, 1281 (1999)
- 50.39 H. Dai, N. Franklin, J. Han: *Appl. Phys. Lett.* **73**, 1508 (1998)
- 50.40 R. H. Baughman, C. Cui, A. A. Zakhidov, Z. Iqbal, J. N. Barisci, G. M. Spinks, G. G. Wallace, A. Mazzoldi, D. De Rossi, A. G. Rinzler, O. Jaschinski, S. Roth, M. Kertesz: *Science* **284**, 1340 (1999)

51. Magnetic Information-Storage Materials

The purpose of this chapter to review the current status of magnetic materials used in data storage. The emphasis is on magnetic materials used in disk drives and in the emerging technology of the magnetic random-access memory (MRAM). A wide range of magnetic materials is essential for the advance of magnetic recording both for magnetic recording heads and media, including high-magnetization soft-magnetic materials for write heads, new antiferromagnetic alloys with high blocking temperatures and low susceptibility to corrosion for pinning films in giant-magnetoresistive (GMR) sensors and new ferromagnetic alloys with large values of giant magnetoresistance. For magnetic recording media, the advances are in high-magnetization metal alloys with large values of the switching coercivity. A significant limitation to magnetic recording is found to be the superparamagnetic effect and new advances in multilayer ferromagnetic films to reduce the impact of the effect, but also to allow high-density recording have been developed.

Perpendicular recording as compared to longitudinal recording is reviewed and it is shown

51.1	Magnetic Recording Technology	1156
51.1.1	Magnetic Thin Films	1157
51.1.2	The Write Head	1158
51.1.3	Spin Valve Read Head	1162
51.1.4	Longitudinal Recording Media	1169
51.1.5	Perpendicular Recording	1175
51.2	Magnetic Random-Access Memory	1185
51.2.1	Tunneling Magnetoresistive Heads	1188
51.3	Extraordinary Magnetoresistance (EMR) .	1189
51.4	Summary	1189
	References	1189

that this technology will be the dominant recording technology in the future. The MRAM device uses some of the same materials as used in the GMR sensor, but the key technology is the magnetic tunneling junction in which soft-magnetic films are coupled by a thin insulating film and conduction is by quantum-mechanical tunneling. The status of the MRAM technology and some of the key problems are reviewed.

The purpose of this chapter is to summarize the status of magnetic materials used in high-capacity disk drives and magnetic-semiconductor memory devices. The technology of disk drives is selected since these devices have experienced the largest increase in data capacity over time and this has made disk drives the pre-eminent storage system for digital data. To illustrate this point, consider Fig. 51.1, which is a plot of the areal density (number of data tracks per inch times the number of bits per inch recorded on a track) for disk drives over time [51.1]. The increase in areal density is more than 100% per year up to about 2002, when it reduced to about 20–30%. At the time of writing the areal density being shipped in disk drives is close to 130 Gb/in². The total data capacity of a disk is approximately the areal density times the recording area, which depends on the size of the disk [2.5 and 3.5 inches (64 and 90 mm)] diameter

being the most common. Many technologies have contributed to this rapid increase in areal density, including advances in the technology of *flying* heads with reduced spacing to the disk surface, data codes and error detection and correction, advanced servo control systems for accurate control of magnetic recording heads on data tracks, and improvements in the mechanical structures comprising a disk drive, including advances in motors used to drive the disks. However, this paper discusses only the fundamental technology associated with digital magnetic recording, including the devices used to record and read back the recorded data and the media on which the data is recorded. The discussion is also restricted to the materials and not to any of the mechanical structures associated with the recording heads or disks. The discussion covers conventional longitudinal magnetic recording as used in present (and past) disk drives

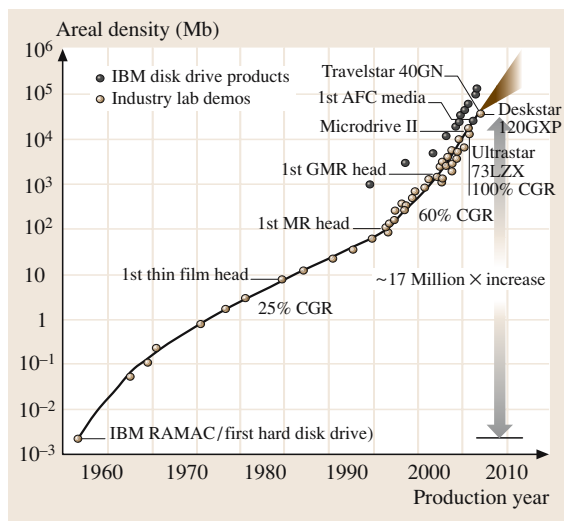


Fig. 51.1 IBM areal density perspective. (After [51.1])

and the emerging technology of perpendicular magnetic recording, in which the remanent magnetization is per-

pendicular to the disk surface, and which reduces the impact of the superparamagnetic limit (to be discussed). We will not discuss some variants of longitudinal magnetic recording, e.g., thermally assisted recording and patterned media.

Recently there has been a significant new emerging technology for fast memory devices – magnetic random-access memory (MRAM). The MRAM device is a possible replacement for the familiar semiconductor memories used in modern computers – dynamic and static random-access memory (DRAM and SRAM). The MRAM technology combines a magnetic storage technology together with metal-oxide semiconductor (MOS) devices to result in fast and high-density data memory devices. The technology on which the magnetic part of MRAM is based is an extension of the technology used in magnetic recording devices – the magnetic tunneling junction (MTJ). The MRAM technology will also be discussed in this paper. Parts of this chapter have been previously published in the *Journal of Materials Science: Materials in Electronics* [51.2].

51.1 Magnetic Recording Technology

The technology of magnetic recording was one hundred years old in 1999 [51.3]. The fundamental concept of magnetic recording is to use a magnetic structure (the *write head*) driven by current which represents the data to be recorded to generate a magnetic field which can change the state of the magnetization in a closely spaced magnetic recording medium, which in the earliest realization was magnetic wire, and today is either

the familiar magnetic tape or a magnetic layer on a rigid disk substrate. The data are recovered by the generation of an output voltage in the *read head* by sensing the magnetization in the recording medium, e.g., by Faraday's law ($V = N \frac{d\Phi}{dt}$), where N is the number of turns on the read head and Φ is the magnetic flux coupled to the read head from the media. The magnetic recording system to be discussed here is that used to store digital data, in which case the current supplied to the write head is in the form of pulses encoded to represent the digital data (1s or 0s) [51.4, 5]. In the case of disk drives the

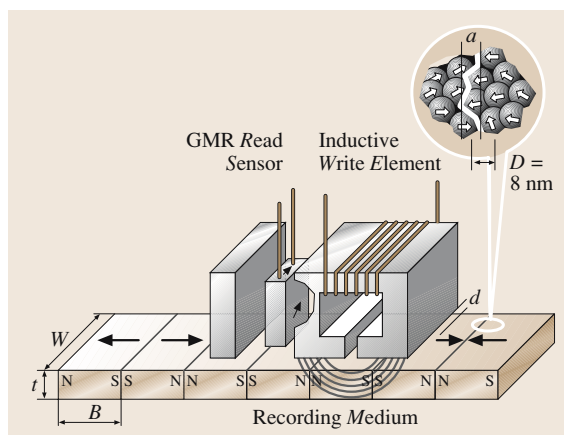


Fig. 51.2 Schematic illustration of a longitudinal magnetic recording system showing a giant magnetoresistive (GMR) read sensor, write element and the recording medium. The parameters are: t the recording medium thickness (δ is also used), W the width of the recorded track, B the length of the recorded magnetization or bit size and d the magnetic spacing (includes overcoats on the disk surface and undercoats on the slider surface). Shown in the *inset* is the transition region between the two directions of the magnetization; D is the media grain size and a is the transition parameter that characterizes the length of the transition region (length $l = \pi a$). (After [51.1])

write and read heads are separate thin-film structures deposited on the back of a mechanical slider, which uses a hydrodynamic air bearing to fly over the surface of the disk [51.5]. Figure 51.2 is a schematic of a modern digital magnetic recording system. The recording (*write*) and read elements are shown together with the magnetic recording surface which in modern disk drive technology is a thin metallic film of a cobalt alloy (to be discussed). The digital data are recorded in the magnetic film as *transitions* between the two possible states of the magnetization (pointing to the *left* or *right*) and with the width approximately equal to the width of the write head and the width of a data track. The transition region between the oppositely directed directions of the magnetization is similar to that between magnetic domains and has a length l . We will discuss the different parts of the recording system starting with the write head and including the read head which in Fig. 51.2 is a giant-magnetoresistive sensor. First, we will discuss some aspects of thin magnetic films that are relevant to magnetic recording components. All magnetic recording components used in disk drives today are fabricated from thin films to allow mass production and for technology reasons, as we will discuss.

51.1.1 Magnetic Thin Films

Induced Uniaxial Anisotropy

In most applications of soft-magnetic films a uniaxial anisotropy in the plane of the film is required. One way of introducing this uniaxial anisotropy is to induce it by applying a magnetic field in the plane of the film and the effect is referred to as magnetic annealing. The anisotropy can be induced during deposition of the film or induced by a subsequent annealing step, using temperature and a magnetic field to modify the anisotropy. The mechanism of the induced anisotropy is that of directional order, in which atomic pairs in an alloy tend to align with the local magnetization. In nickel-iron films the atomic pairs are the iron atoms and the induced uniaxial anisotropy energy density is typically $K_u \cong 1\text{--}3 \text{ kerg/cm}^3 (0.1\text{--}0.3 \text{ kJ/m}^3)$, where the uniaxial energy density is

$$E_k = K_u \sin^2 \theta, \quad (51.1)$$

where θ is the angle of the magnetization with respect to the direction of the induced anisotropy. To induce the anisotropy it is necessary to saturate the magnetization of the film with a small magnetic field, typically 50–100 Oe (4000–8000 A/m), since it is the magnetization not the magnetic field that is responsible for the

magnetic annealing. If the anisotropy energy density is positive the energy is minimum is along the direction of the anisotropy, which is referred to as an *easy* axis.

A second method of introducing uniaxial anisotropy in thin films is by stress. If the magnetostriction constant is isotropic with value λ_s for a polycrystalline material then the magnetoelastic energy density is given by

$$E_{me} = \frac{3}{2} \lambda_s \sigma \sin^2 \theta. \quad (51.2)$$

The angle θ is the angle of the magnetization in the film plane with respect to the direction of the stress (σ). If a sample is unmagnetized it will strain by λ_s as the sample is magnetized to saturation. A typical value of λ_s is $\approx 10^{-6}$. If the magnetostriction constant and the stress are both positive the direction for the minimum energy and hence the easy axis is for $\theta = 90^\circ$. The magnetization, permeability, crystalline anisotropy constant and magnetostriction constant of nickel-iron alloys $\text{Ni}_x\text{Fe}_{(1-x)}$ are shown in Fig. 51.3 [51.6]. The maximum permeability is for the Permalloy composition $\text{Ni}_{80}\text{Fe}_{20}$. Increasing the atomic percentage of iron over that of Permalloy results in increased magnetization since the magnetic moment for iron is $2.2 \mu_B$ (Bohr magnetons) per atom in the metal, while the value for nickel is $0.6 \mu_B$ per atom in the metal.

Hysteresis in Soft-Magnetic Films

If we impose an external magnetic field H_e along the direction of the induced anisotropy (easy axis) of the thin film ($\theta = 0^\circ$) the total energy density is given by

$$E_k = K_{\text{eff}} \sin^2 \theta - MH_e \cos \theta.$$

K_{eff} is the sum of the uniaxial crystalline anisotropy (K_u) and the stress anisotropy ($3/2 \lambda_s \sigma$). The second term is the magnetostatic energy. The equilibrium angle for the magnetization is given by minimizing the total energy,

$$\frac{dE}{d\theta} = 2K_{\text{eff}} \sin \theta \cos \theta + MH_e \sin \theta = 0.$$

The solution $\sin \theta = 0$ implies that the magnetization lies at either $\theta = 0$ or 180° . However, the energy must be a minimum for stability and the second derivative changes sign at a boundary between stability and instability

$$\frac{d^2E}{d\theta^2} = 2K_{\text{eff}}(\cos^2 \theta - \sin^2 \theta) + MH_{e,\text{crit}} \cos \theta = 0.$$

The magnetic field which just satisfies this condition is

$$H_{e,\text{crit}} = -H_c = -\frac{2K_{\text{eff}}}{M}. \quad (51.3)$$

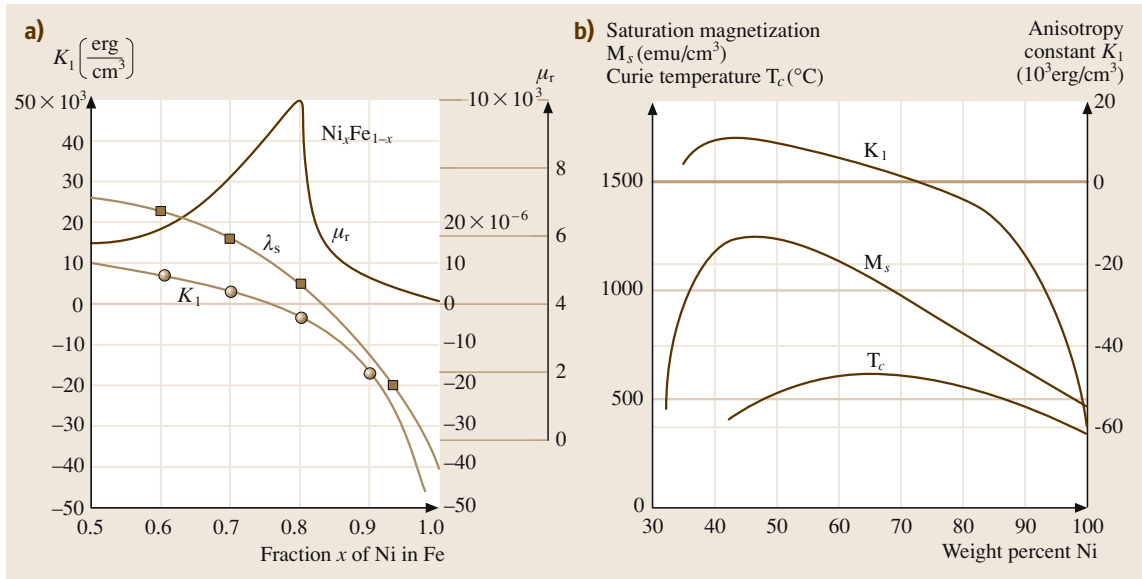


Fig. 51.3 (a) Magnetic properties of $\text{Ni}_x\text{Fe}_{1-x}$ alloys, including initial permeability, crystalline anisotropy and magnetostriction. (b) Magnetization, Curie temperature and crystalline anisotropy of $\text{Ni}_x\text{Fe}_{1-x}$ alloys. (After [51.6])

The critical magnetic field is the coercive field (H_c). A curve of the magnetization versus applied magnetic field for this case is shown in Fig. 51.4 (curve labeled “easy”), which was measured on a thin film of FeCoNi (to be discussed) [51.7]. The magnetization curve exhibits hysteresis and in the FeCoNi alloy the coercive field is small and the material is magnetically *soft*. Hysteresis is responsible for the use of magnetic materials as digital memory devices since the *up* or positive state

of the magnetization could represent a binary 1 and the *down* or negative state could represent a binary 0 and these states can be stable even in zero applied field ($H_e = 0$). The above analysis also applies to magnetically hard films, for which the coercive field is large. We will discuss magnetic recording media and here the coercive field is measured in units of kOe.

With the magnetic field applied perpendicular to the easy axis, along the hard axis, the energy density is given by

$$E_k = K_{\text{eff}} \sin^2 \theta - MH_e \sin \theta.$$

The derivative of the total energy is given by

$$\frac{dE}{d\theta} = 2K_{\text{eff}} \sin \theta \cos \theta - MH_h \cos \theta = 0$$

resulting in the two solutions

1. $\cos \theta = 0$, or $\theta = 90^\circ$ or 270° ;
2. $\sin \theta = \frac{H_h}{2K_{\text{us}}/M}$.

The solution to this case is also shown in Fig. 51.4 (curve labeled “hard”) and is seen not to exhibit hysteresis. This is the *hard-axis* magnetization curve and the magnetic field where the magnetization saturates is known as the anisotropy field H_k . The hard-axis magnetization curve is used extensively for magnetic recording heads.

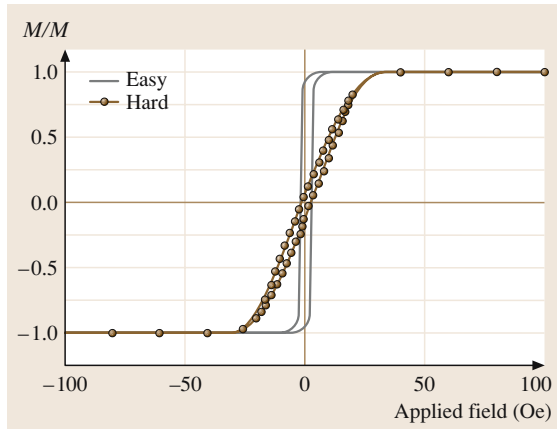


Fig. 51.4 Easy and hard magnetization curves of an electroplated FeCoNi alloy film. (After [51.7])

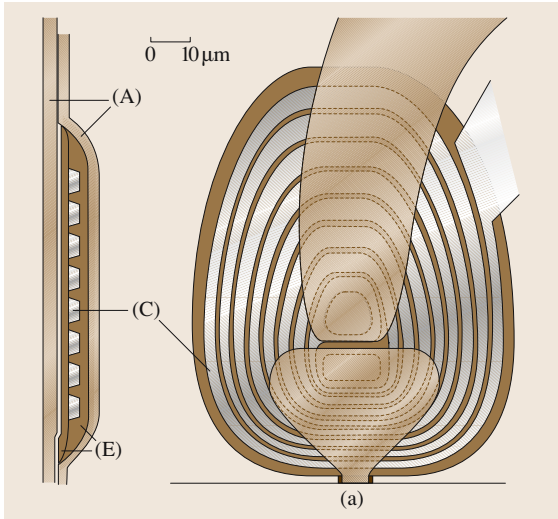


Fig. 51.5 Thin film head with eight turns showing the poles and yoke structure. A is Permalloy, C is copper and E is baked photoresist. (After [51.13])

51.1.2 The Write Head

The write head is formed from thin films of soft ferromagnetic alloys patterned in the form of a magnetic yoke. A schematic of an eight-turn thin-film head is shown as Fig. 51.5 [51.13]. Heads in present disk drives typically have fewer than five turns to reduce inductance. The cross-hatched regions are the soft ferromagnetic material and the leads which carry the write current are shown on the right-hand side of the top view. Note that the widths of the copper alloy turns are increased near the back region of the head to reduce resistance. The heads are fabricated using electroplating and the substrate (not shown) is a hard ceramic that can be machined to form the slider [51.5]. Electrical current in the form of pulses is coupled to the yoke to generate a magnetic field at

the gap. The coil is insulated from the metallic magnetic yoke by layers of baked polymer photoresist. The alloy that has been used most frequently in the past for the magnetic films in the write head is $\text{Ni}_{80}\text{Fe}_{20}$ – Permalloy, which can be deposited in thin films using electroplating. Electroplating is discussed in Comstock [51.4] and in more detail by Andricacos and Romankiw [51.14]. The maximum coercivity ($H_{c,\max}$) of the recording medium that can be recorded by the write head is given by (51.4) (in cgs units) [51.4]

$$H_{c,\max} = 0.20 \times 4\pi M_s \tan^{-1} \left(\frac{g/2}{d + \delta} \right), \quad (51.4)$$

where $4\pi M_s$ (Gauss) is the saturation flux density of the material used in the recording or write head, g is the gap length of the write head, d is the magnetic spacing, which is the spacing of the head to the recording magnetic film (including any overcoat films on the recording heads and recording medium) and δ is the thickness of the recording film (shown as t in Fig. 51.2). To increase the linear density (transitions along a recorded track), it is necessary to increase the coercive field of the recording media. This is the case since the length of a transition between the two states of the magnetization is given by (51.5) (in cgs units) [51.15] and [51.4].

$$l = \pi a = \pi \left(\frac{d(1 - S^*)}{\pi Q} \right) + \left[\left(\frac{d(1 - S^*)}{\pi Q} \right)^2 + \left(\frac{2M_r \delta}{H_c} \right) \left(\frac{2d}{Qr} \right) \right]^{1/2}, \quad (51.5)$$

where $S^* = 1 - \frac{M_r/H_c}{dM/dH(H=-H_c)}$ quantifies the slope of the hysteresis curve of the recording medium at the coercive field and $Q = \frac{dH_x/dx}{H_c/d}$; Q is the normalized slope of the horizontal component of the write-head magnetic field at the value $H = H_c$. A typical value of Q is 0.7.

Table 51.1 Properties of materials used in magnetic recording heads

Material	B_s (kG) ($\text{T} \times 10^{-1}$)	μ_r	ρ ($\mu\Omega\text{cm}$)	H_c (Oe) (A/m)	References
$\text{Ni}_{80}\text{Fe}_{20}$ (Permalloy)	10	1500–3000	20	0.1–0.5 (7.95–39.78)	[51.6]
Sendust (FeSiAl)	10	8000	85	0.25 (19.9)	[51.4]
$\text{Ni}_{45}\text{Fe}_{55}$	16	1700	48	0.3 (23.87)	[51.8]
$\text{Co}_{48}\text{Ni}_{22}\text{Fe}_{30}$ electroplated with pulsed plating	21.5	1000	–	≈ 1 (79.6)	[51.9]
$\text{Co}_{56}\text{Ni}_{13}\text{Fe}_{31}$ (C) electroplated with carbon impurity	19	–	≈ 70	≈ 2 (159.2)	[51.10]
$\text{FeAl}(2\%\text{Al})\text{N}$ (N/Fe 6.6%) sputtered	20	2450	38	3.8 (302.4)	[51.11]
FeTiN sputtered	18–21	–	–	≈ 5 (397.9)	[51.12]

M_r (emu/cm³) is the remanent magnetization of the recording medium with thickness δ .

From (51.4) and (51.5) it is seen that for reduced transition length (and hence higher linear density) it is necessary to increase the coercive field (H_c) of the recording medium but to do so requires a concomitant increase in the magnetization of the write head (51.4). Materials with magnetization larger than Permalloy [$4\pi M_s = 10$ kG ($M_s = 1$ T)] and which are magnetically soft (large permeability) are listed in Table 51.1. The most common material used in present disk drives is electroplated Ni₄₅Fe₅₀, with a saturation magnetization of $4\pi M_s = 16$ kG ($M_s = 1.6$ T) [51.8]. Ternary alloys of Ni, Fe and Co also can be electroplated with saturation magnetization of $4\pi M_s \geq 20$ kG ($M_s = 2$ T). The composition region that results in the largest value of saturation magnetization but also large permeability is: Co_{48–56}Fe_{30–31}Ni_{22–13} (Table 51.1). A useful guide to the selection of the composition of the ternary FeCoNi alloys is the composition diagram shown as Fig. 51.6 [51.6]. The data for Fig. 51.6 were accu-

mulated in bulk materials. In a study of electroplating FeCoNi alloys for write heads [51.7] care was taken to avoid any impurities in the plating bath and the results showed the possibility of achieving $4\pi M_s$ values in the range of 21–21.5 kG and $H_c < 2$ Oe in the same region of the composition diagram (approximately Fe₃₅Co₄₅Ni₂₀). However, experiments in a larger cell in conditions closer to those in a manufacturing plating system and with higher Fe content and an organic additive ascorbic acid used to reduce the formation of Fe³⁺ showed even larger values of $4\pi M_s$ (> 2.2 kG) but with increased coercivity (> 8 Oe) [51.16]. The composition of ferromagnetic alloys can be selected to result in zero values for the magnetostriction constant and the Permalloy composition of nickel–iron alloys (Ni₈₀Fe₂₀) has a zero magnetostriction constant (Fig. 51.3). The magnetostriction constant is frequently chosen to be small and negative so that the magnetoelastic energy defines an easy axis at right angles to the direction of the stress [$\theta = 90^\circ$ in (51.2)]. In the case of thin-film write heads the stress is along the axis of the poles (upward in the

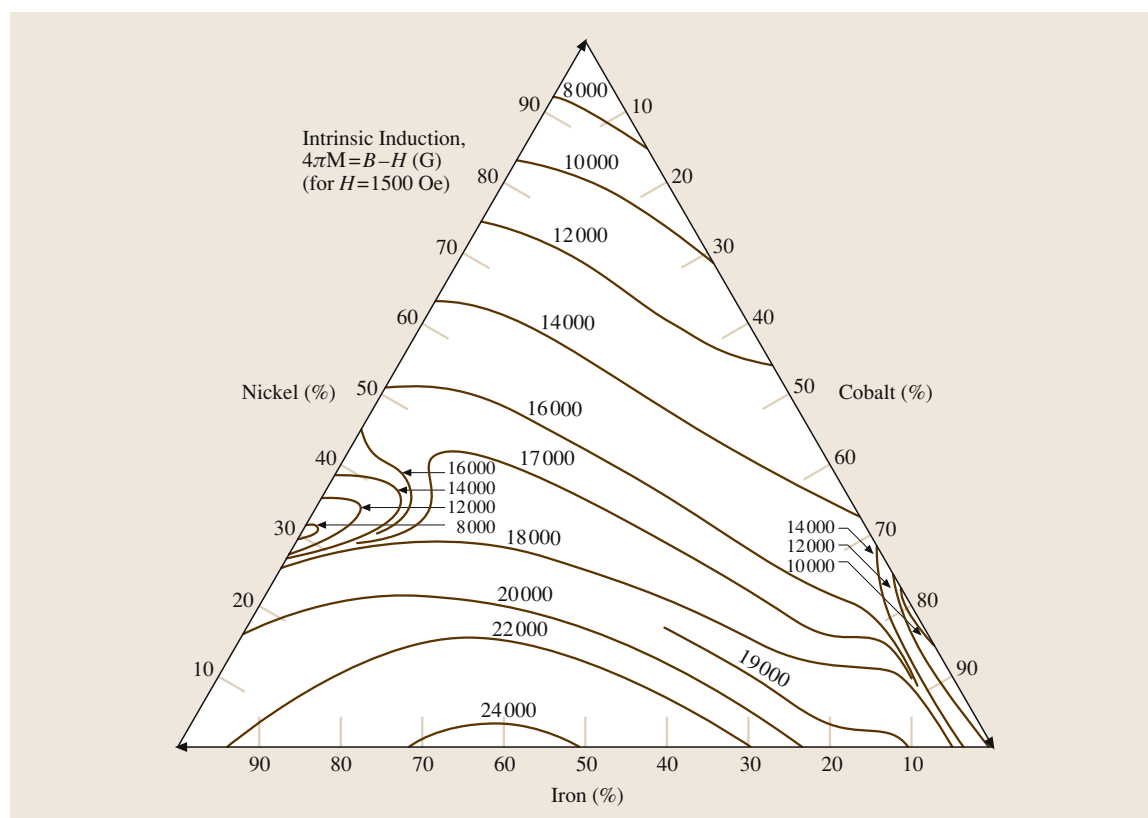


Fig. 51.6 FeCoNi saturation magnetization contours. (After [51.6])

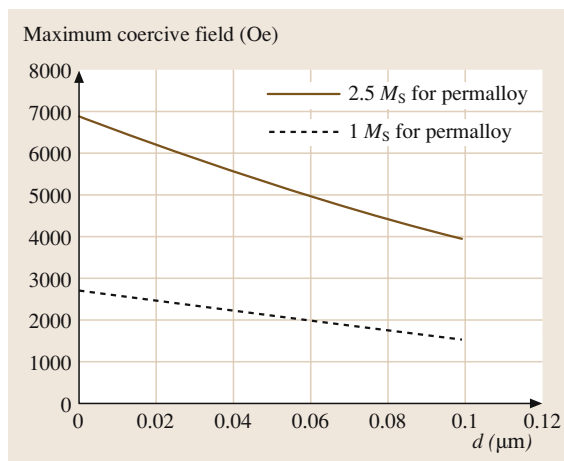


Fig. 51.7 Maximum coercivity of recording media versus magnetic spacing (d) for write heads fabricated from Permalloy and from a magnetic material with saturation magnetization $2.5\times$ Permalloy

patterned yoke in Fig. 51.5) and with negative magnetostriction constant the easy axis is across the pole width, which means that the direction of flux flow is along the hard axis with the desired linear relation between the magnetization and the magnetic field (Fig. 51.4). In the case of the ternary alloys of FeCoNi an alloy composition that results in zero magnetostriction is $\text{Fe}_{13}\text{Co}_{31}\text{Ni}_{56}$, in which the nickel content is too high to be a useful write-head material [51.7].

Alloys of FeAl (2% Al) [51.11] and FeTi ($\approx 2\%$ Ti) [51.12] sputtered using a mixed N and Ar working gas offer desirable magnetic properties, as shown in Table 51.1, but have not replaced the electroplated alloys because of increased costs of sputtering. A material of interest for even higher values of saturation magnetization is the α'' phase of Fe_{16}N_2 , which has a potential of a saturation magnetization of 29 kG; however, this phase is metastable [51.18]. The actual values of saturation magnetization with this phase in thin-film form was 20 kG and required annealing at 200°C [51.18].

To record on recording media with increased coercivity is not the only issue with the magnetic materials used in write heads. It is also important that the write heads have high efficiency. Efficiency (η) in this case is defined as the ratio

$$\eta = \frac{H_g g}{N_w I}, \quad (51.6)$$

where H_g is the value of the magnetic field in the gap of the write head and I is the amplitude of the write

current pulse. High efficiency is important to allow write-current amplitudes that are easily supplied from integrated circuits. For high efficiency it is necessary that a large percentage of the magnetomotive force ($N_w I$) results in a magnetic field across the gap. However, the permeability of the yoke influences the efficiency

$$\eta = \frac{g/A_g}{g/A_g + l_c/\mu_r A_c}, \quad (51.7)$$

where A_g and A_c are the area of the gap and core, respectively; l_c is the length of the core and μ_r is the relative permeability of the core or yoke material. As shown in Fig. 51.3, the relative permeability (μ_r) of nickel-iron alloys drops for compositions different from the Permalloy composition. This decrease in the permeability of nickel-iron alloys with composition different from that for Permalloy is the direct result of the increase in crystalline anisotropy and magnetostriction, as shown in Fig. 51.3. The lower permeability of the FeCoNi alloys can be obviated in a design of the write head in which the high-magnetization films are used only in the gap region in a design referred to as a *stitched head*.

The values of coercivity of the recording medium (H_c) that can be recorded with recording heads with the saturation magnetization of Permalloy [$4\pi M_s = 10\text{ kG}$ ($M_s = 1\text{ T}$)] and materials with saturation magnetization 2.5 times that for Permalloy [$4\pi M_s = 25\text{ kG}$ ($M_s = 2.5\text{ T}$)] are shown versus the magnetic spacing (d) in Fig. 51.7. The magnetic spacing for recent technology demonstrations has reached a value of 15 nm (carbon overcoat for head – 3 nm, pole tip recession tolerance – 3 nm, flying height – 5 nm, and media

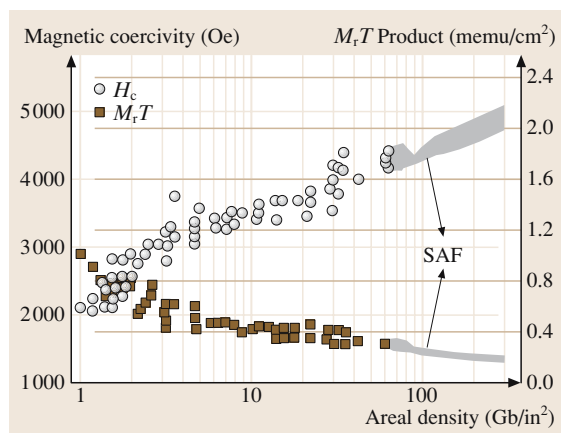


Fig. 51.8 Trend in coercive field (H_c) and media flux ($M_r T$) versus areal density in Gb/in^2 for disk drives. (After [51.17])

carbon overcoat—4 nm). The trend in coercivity (H_c) and flux ($M_r\delta$ or M_rT) versus areal density in commercial disk media is shown as Fig. 51.8 [51.17]. We will discuss synthetic antiferromagnetic media (SAF) media later.

With the increased linear density associated with increased areal density in disk drives and with increases in the rotation speed of the disks, the rate at which data are recorded (and read back) has increased over time. This increase in data rate leads to a requirement that the magnetic fields in the gap of a write head switch in ever decreasing times. The switching time of thin magnetic films as used in write heads is limited by three factors: first, the fundamental switching time resulting from the fact that magnetic moments are always associated with angular momentum that cannot be changed instantaneously; second, the eddy currents associated with the changing magnetic fields both decrease the magnetic fields and slow down the switching speed, and third the inductance of the coil [51.21]. Data rates depend on the form factor of the disks and with high-end 3.5-in drives the data rates are larger than 50 Mb/s, while with 2.5-in drives the data rates are roughly one-half as large.

51.1.3 Spin Valve Read Head

The technology that has evolved for sensing or reading the flux emanating from transitions in thin-film disk

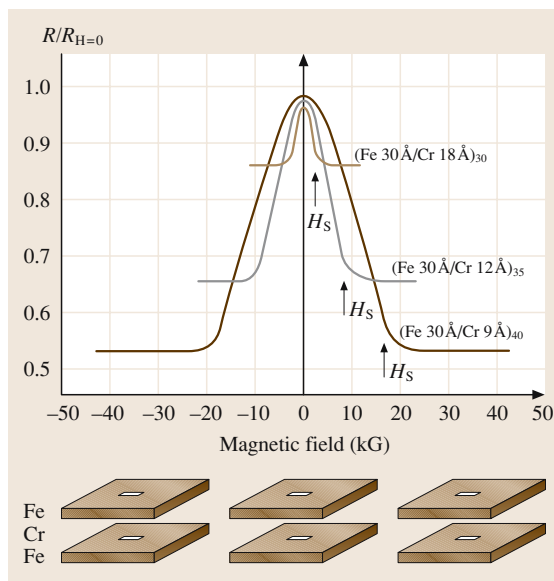


Fig. 51.9 The magnetoresistance of three $(\text{Fe}/\text{Cr})_n$ superlattices at 4.2 K. The thickness of the layers in angstroms is shown. (After [51.19])

media is the spin valve. This technology is based on the giant-magnetoresistive (GMR) effect discovered in France and published in 1988 [51.19]. GMR results when thin ferromagnetic films are coupled by an even thinner spacer film of a nonmagnetic transition metal. A curve of resistance normalized to that with no magnetic field applied versus magnetic field for arrays of ferromagnetic films of Fe with thickness of 30 Å (3 nm) coupled by Cr with thickness varying from 9–18 Å (0.9–1.8 nm) is shown in Fig. 51.9 [51.19]. The number subscripted to the parenthesis characterizing the magnetic and transition-metal films is the number of such films, which are referred to as *superlattices*. With no applied magnetic field, the two films were antiferromagnetically coupled; that is, the magnetizations in the two films are equal in magnitude and oppositely directed. The application of a large in-plane magnetic

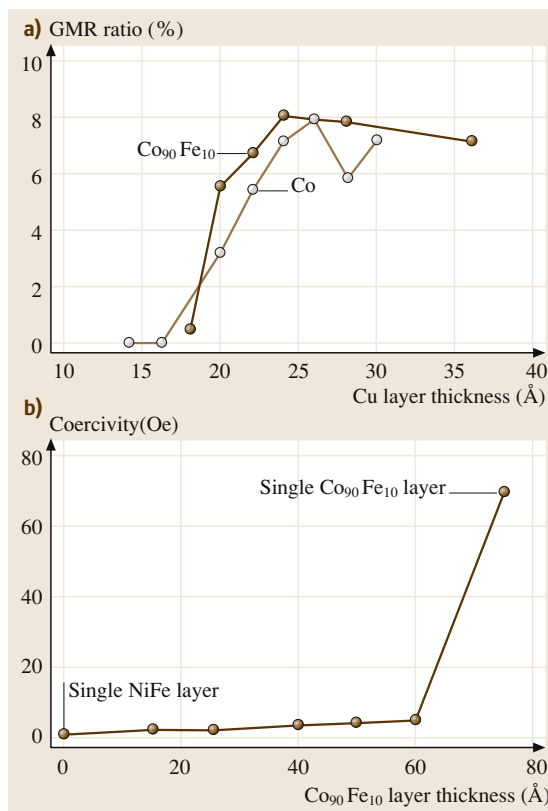


Fig. 51.10 (a) Comparison of GMR ratio ($\Delta\rho/\rho$) as a function of copper layer thickness in $\text{Co}_{90}\text{Fe}_{10}$ and Co films (see text). (b) Coercivity as a function of $\text{Co}_{90}\text{Fe}_{10}$ layer thickness in Ta (50 Å)/ $[\text{Ni}_{80}\text{Fe}_{20}/\text{Co}_{90}\text{Fe}_{10}]$ (75-Å) films. (After [51.20])

field rotated the magnetization of the coupled films to a parallel configuration, reducing the resistance (to be discussed). The current used to measure the magnetoresistance – the change in resistance with applied magnetic field – is applied in the plane of the films. This orientation of current is referred to as current-in-plane (CIP). Later we will discuss an alternative orientation of the current – current-perpendicular-to-plane (CPP). The initial discovery of **GMR** was done with a temperature of 4.2 K and the magnetic fields applied were measured in kOe ($10^6/4\pi$ A/m). The large change in resistance with magnetic field was in contrast to that observed in the anisotropic magnetoresistance (AMR) effect, which is the change in resistance observed when the magnetization in a *single* ferromagnetic film is rotated from an easy toward a hard axis. This latter technology – for which the resistance change with Permalloy was of the order of 2–3% – was being used in read heads during the time of research on the **GMR** effect. By the early 1990s it was found possible to observe **GMR** in other film materials with increased spacer film thickness at 300 K and with magnetic fields of tens of Oersteds. For example, the **GMR** ratio for two coupled films of an alloy of cobalt and iron ($\text{Co}_{90}\text{Fe}_{10}$ for which the magnetostriction constant is zero) with a variable thickness of the spacing layer of copper is shown in Fig. 51.10a [51.20]. It is seen that the **GMR** ratio is considerably reduced from that for the Fe/Cr superlattices at 4.2 K (Fig. 51.8) but the ratio is larger than that for the AMR effect. The discovery of the **GMR** effect led to the invention of the *spin valve read head*, in which one of two ferromagnetic films coupled by a thin transition-metal film is maintained constant in direction (the *pinned* film) and the orientation of the other film (the *free* film) is allowed to rotate (in the plane of the magnetic film), resulting in a decrease in resistance [51.22]. We will discuss the technology of pinning later. By free, it is meant that the magnetization in the film is easily rotated, e.g., by the magnetic field arising from transitions in the recording medium. It is surprising that cobalt is found to be useful in this application since cobalt is magnetically hard (pure cobalt metal has a high coercivity and is used in recording surfaces) and one of the requirements for the **GMR** effect is that the free film have high permeability and high efficiency (51.6). The reason for this result is shown in Fig. 51.10b, where the coercivity of a free film consisting of two films: $\text{Ni}_{80}\text{Fe}_{20}$ [60 Å (6 nm)] and $\text{Co}_{90}\text{Fe}_{10}$ is plotted versus the thickness of the $\text{Co}_{90}\text{Fe}_{10}$ film [51.20]. Even though the cobalt alloy would have a large coercivity in the bulk, the coercivity for thin films, together with the underlying $\text{Ni}_{80}\text{Fe}_{20}$ film, is low

for thicknesses less than about 60 Å. The advantage of the cobalt-iron film is the increased **GMR** ratio over $\text{Ni}_{80}\text{Fe}_{10}$. In order to increase the **GMR** ratio even more it has been found advantageous to deposit an interfacial nanolayer of cobalt metal on one or both sides of the free film [51.23]. We will learn in a discussion of the origin of **GMR** that cobalt has the potential for large **GMR** ratios.

A schematic of the spin valve read head is shown as Fig. 51.11 [51.22]. The magnetization of the free film is shown rotated to an angle θ_1 and the pinned layer is shown with a fixed angle of θ_2 . The pinning effect is due to the coupling of the pinned film to an underlying anti-ferromagnetic film as we will discuss. The nonmagnetic spacing layer separates the two ferromagnetic films. The maximum output voltage for the spin valve with no external resistance (and no reduction in the flux due to shielding (to be discussed) is

$$\Delta V = \eta J W_r \Delta \rho (1 - \sin \Delta \theta_1) / 2, \quad (51.8)$$

where η is the efficiency of the read head, which accounts for the fact that the angle θ_1 is not uniform over the height of the free film, J is the density of the current flowing through the films, W_r is the width of the read head (always less than the width of the written track) and $\Delta \rho$ is the change in resistivity of the stack of films with a change of θ_1 from -90° to $+90^\circ$, again with θ_2 fixed. The factor of 2 in the denominator is the result of the actual rotation of θ_1 being limited to 0° to 90° . The device senses magnetic field in the plane of the free film resulting from transitions in the recorded media ($H_{y,\text{disk}}$)

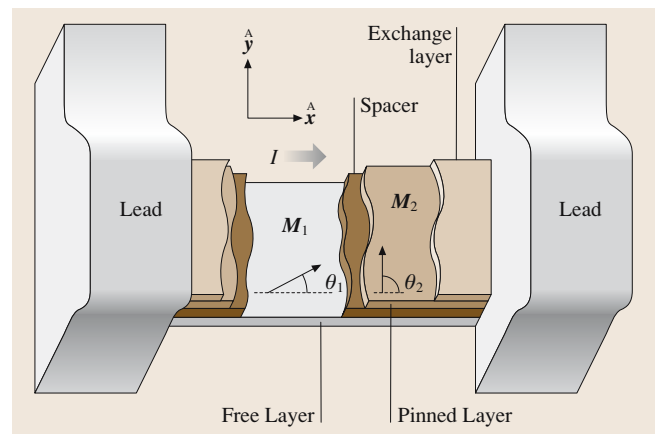


Fig. 51.11 Schematic of spin valve read element. The magnetization of the free layer is shown at an angle θ_1 to the easy axis (and direction of the current) and the magnetization of the pinned layer is fixed at 90° to this axis. (After [51.22])

through the rotation of the magnetization in the free film

$$\sin \theta_1 = \frac{H_{y,\text{disk}}}{H_k}, \quad (51.9)$$

where H_k is the anisotropy field of the free film (see analysis of the hard-axis magnetization curve in Sect. 51.1.1). In the absence of any other magnetic fields, the magnetization will lie along the easy axis. The hard axis is perpendicular to the easy axis. The voltage is measured across the lead terminals using a sense current and the width of the track is just the spacing between the leads. A typical value of peak output voltage per unit of track width for presently used spin valves is 15 mV/ μm . Spin valve read heads are shielded on each side of the stack of films to realize a narrow pulse width. The total width between the shields is G and the thickness of the stack of films comprising the spin valve is t . The read pulse is characterized by the amplitude but also by the width at one-half amplitude (PW_{50} in distance units) (discussed in [51.4])

$$PW_{50} \approx \sqrt{\frac{1}{2} (G^2 + t^2) + 4 (d + a + \delta/2)^2}, \quad (51.10)$$

where G is the total spacing between the shields and δ is the thickness of the recording layer.

The peak voltage amplitude with the shields with no external resistance can be approximated by

$$V_{\text{GMR}}(\text{peak}) = \frac{\eta I \Delta R_{\text{max}} M_r^{\text{rec}} \delta}{\pi t M_s^{\text{GMR}}} \left(\frac{(g+t)}{g} \right) \times 2 \tan^{-1} \frac{g}{2(d+a)},$$

$$I \Delta R_{\text{max}} = J W_{\text{read}} R_{\text{square}} t_{\text{sv}} \frac{\Delta R_{\text{max}}}{R_0}, \quad (51.11)$$

where η is the read efficiency, defined as the fraction of the flux sensed by the head to the total flux, M_r^{rec} is the magnetization of the recording layer, M_s^{GMR} the saturation magnetization of the free layer, g is the half-gap between the shields, R_{square} is the resistance per square of the stack of films comprising the spin valve and t_{sv} is the thickness of the free layer (modified from [51.24]).

Pinning with an Antiferromagnetic Film

For a spin valve to function, the pinned ferromagnetic film must have its magnetization perpendicular to the recorded track, as shown in Fig. 51.11 (the vertical magnetic field from the transitions in the disk is in the y -direction). This orientation of the magnetization in

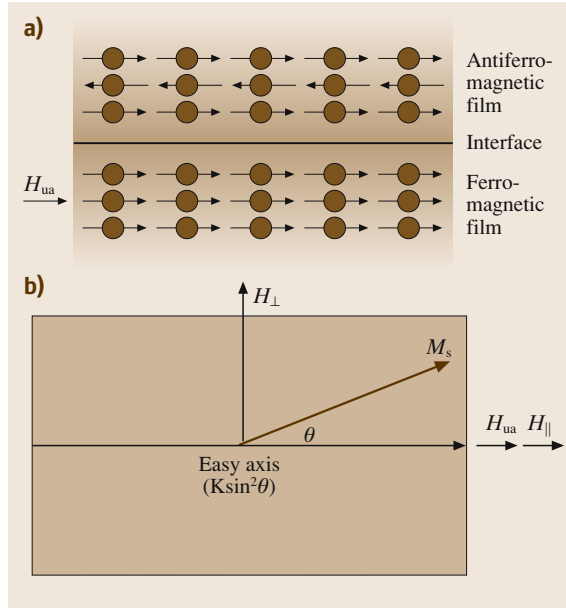


Fig. 51.12 (a) Schematic of the coupling of an antiferromagnetic to a ferromagnetic film. (b) Geometry of the ferromagnetic film showing the induced easy-axis field, H_{ua} , the external magnetic fields H_{\parallel} and H_{\perp} and the magnetization of the film at an angle θ to the easy axis. (After [51.4])

the pinned film is achieved by coupling the ferromagnetic film to an antiferromagnetic film and the coupling results from the layer of spins in the antiferromagnetic film adjacent to the top layer of spins in the ferromagnetic film and is characterized by a coupling or exchange field of value H_{ua} . The geometry for the coupling is shown in Fig. 51.12a and the ferromagnetic film geometry with a magnetic field along the easy axis (H_{\parallel}) and the hard axis (H_{\perp}) as Fig. 51.12b. The coupling results in shifting the easy axis hysteresis curve (M_{\parallel}/M_s versus H_{\parallel}) of the ferromagnetic film by an amount H_{ua} , as shown in Fig. 51.13 [51.4]. The magnetization of the pinned film is held constant at $-M_s$ over the range of magnetic fields from large negative fields to $H_{\text{ua}} + H_k$. Note that the magnetization could be held at $+M_s$ by reversing the direction of the lower layer of spins in the antiferromagnetic film. The coupling field depends on the temperature and vanishes at the *blocking* temperature. The direction of the coupling is established during an annealing step by heating the structure above the blocking temperature, where the film is paramagnetic, in the presence of a magnetic field and then cooling the film below the blocking temperature. The choice of

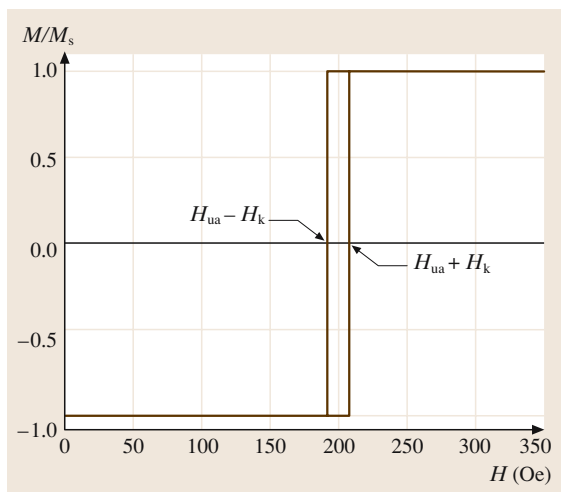


Fig. 51.13 Magnetization curve with the magnetic field along the easy axis ($M_{||}/M_s$ versus $H_{||}$) for coupled antiferromagnetic and ferromagnetic films. (After [51.4])

antiferromagnetic films to be used in this application is extensive as shown in Table 51.2 [51.25–28]. The parameters listed in Table 51.2 are the strength of the coupling field (H_{ua}), the blocking temperature and the susceptibility of the antiferromagnetic and ferromagnetic films to corrosion. Corrosion is of concern in recording heads because of the mechanical lapping process used to form the surface on the bottom of the slider on which the air-bearing contour is etched [51.5]. One antiferromagnetic film with particularly favorable corrosion susceptibility is NiO. This film is typically used in a bottom spin valve in which the antiferromagnetic film is closest to the substrate. The other antiferromagnetic films are typically formed on top of the stack of films. Another advantageous film is the alloy of $\text{Pd}_{30}\text{Pt}_{20}\text{Mn}_{50}$ [51.28]. This ternary alloy has a favorable coupling strength H_{ua} but also a high blocking temperature (300 °C) and favorable

corrosion resistance. $\text{Pd}_{30}\text{Pt}_{20}\text{Mn}_{50}$ does require annealing the stack at 220–240 °C. The advantage of using an antiferromagnetic film for the pinning film instead of a permanent magnetic film is that the antiferromagnetic film does not itself respond to any external magnetic field since the film has a zero net magnetization.

Synthetic Antiferromagnets

The magnetic coupling between magnetic films closely spaced by certain nonmagnetic metal films oscillates between antiferromagnetic and ferromagnetic coupling. Using this fact, it is possible to design film structures that have a desired magnitude and sign of the magnetic coupling [51.29]. A schematic of a structure of films that exhibits this variable coupling is shown in Fig. 51.14a. Here the strong antiferromagnetic coupling of Co through Ru to soft $\text{Ni}_{80}\text{Co}_{20}$ is used to pin the magnetization in the adjacent $\text{Ni}_{80}\text{Co}_{20}$ film (F_I) to the direction opposite to the Co film and the thickness of the second layer of Ru can be adjusted to change the sign of the coupling between the two soft-magnetic films (F_I and F_{II}). Ruthenium is selected for the coupling film since it has a strong and oscillating coupling between Co, $\text{Ni}_{80}\text{Fe}_{20}$ and $\text{Ni}_{80}\text{Co}_{20}$. Ruthenium has the hexagonal close-packed (hcp) crystal structure with ($a = 0.271$ nm and $c = 0.428$ nm) while cobalt also has the hcp structure with ($a = 0.250$ nm and $c = 0.410$ nm) and therefore the two metals are expected to grow epitaxially. The experimental coupling constant (J_{12}) for this configuration of layers, where the soft-magnetic film is $\text{Ni}_{80}\text{Co}_{20}$, is shown as Fig. 51.14b. The units of the coupling constant are memu/cm^2 ($\times 10$ A in SI units) (coupling per unit of film area). The coupling constant can be either antiferromagnetic ($J_{12} < 0$) or ferromagnetic ($J_{12} > 0$) and can have a magnitude varying between +25 to -75 memu/cm^2 (+250 to -750 A). By varying the thickness of the Co layer it is possible to adjust the total magnetic field at the layer F_{II} to be zero.

Table 51.2 Comparison of alternative antiferromagnetic films for transverse bias of Permalloy films

Property	NiFe/FeMn [51.25]	NiFe/NiMn [51.25]	NiO/NiFe [51.25]	CoNiO/NiFe [51.26]	NiFe/IrMn [51.26]	NiFe/CrMnPt [51.27]	NiFe/Pd–Pt–Mn [51.28]
Corrosion resistance	Low	High	Very high	Excellent	Moderate	Good	High
Coupling field H_{ua} Oe (kA/m)	≈ 77 (6.12)	≈ 206 (16.4)	≈ 46 (3.66)	≈ 45 (3.58)	60 (4.77)	≈ 350 (27.84)	($\text{Pd}_{0.6}\text{Pt}_{0.4}$) $_{0.5}\text{Mn}_{0.5}$ 80 (6.36)
Coercivity H_{ua} Oe (kA/m)	≈ 6 (0.48)	≈ 116 (9.23)	≈ 35 (2.78)	11 (0.88)	8 (0.64)	≈ 23 (1.82)	16 (1.27)
Blocking temperature (°C)	≈ 150	≥ 450	≈ 200	150	250	320	300
Annealing	Not required	Required	Not required	Not required	Not required	Required	Required

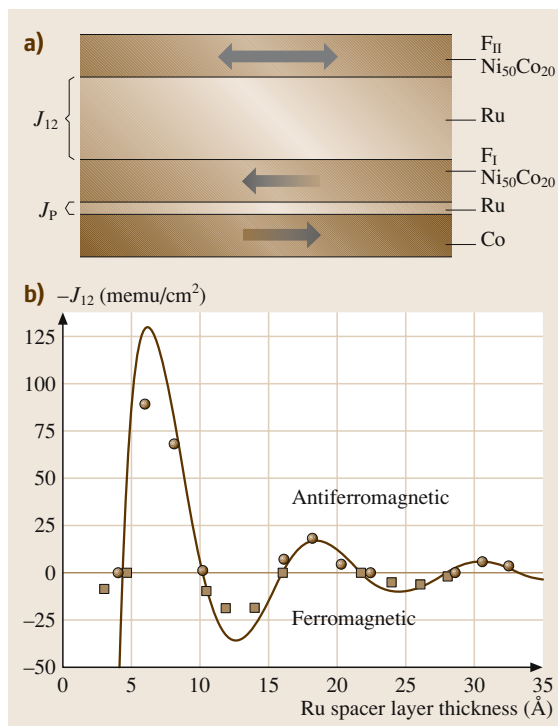


Fig. 51.14 (a) Geometry of ferromagnetic films coupled by thin layers of ruthenium. (b) Magnetic coupling constant per unit of film area between layers F_I and F_{II} of the films shown in (a) versus the thickness of the ruthenium layer. (After [51.29])

An overall spin valve consists of a sequence of thin films such as: MnFe 21 Å/Co 30 Å/Ru 7 Å/Co 30 Å/Cu 25 Å/Co 30 Å/Ni₈₀Fe₂₀ 28 Å, where the first four films (MnFe/Co/Ru/Co) form a *synthetic antiferromagnetic* film since the final Co film of these four films is pinned to a direction opposite to that for the first Co film. This Co film then acts as a pinned film in a spin valve together with the final three films (Cu, Co, and Ni₈₀Fe₂₀). The thickness of ruthenium listed in the sequence of films is optimum for antiferromagnetic coupling between the Co films. The last Co film is a nanolayer, which increases the magnetoresistive coefficient, as we have discussed [51.23]. The advantages of the synthetic antiferromagnet for pinning the pinned film compared with just an antiferromagnetic film (e.g., FeMn, NiO, NiMn or IrMn) are

1. the pinning coupling field is an order-of-magnitude larger, and

2. the magnetostatic coupling field from the pinned layer to the free layer is reduced. The magnetostatic coupling field is reduced since the two Co films in the synthetic antiferromagnet are antiferromagnetically coupled and hence their magnetostatic fields at the free film cancel. The synthetic antiferromagnet contributes to ensuring that the only magnetic field at the free film is that from the transitions in the recording medium.

One last requirement for the proper operation of the spin valve is that there is abutted on the two ends of the structure a permanent magnetic film referred to as the *hard bias* film (discussed in [51.4]). The permanent magnetic films are similar to the cobalt alloy magnetic films used for disk recording surfaces (to be discussed). The purpose of the hard bias is to reduce the presence of magnetic domains in the free film and to reduce Barkhausen jumps of the domain walls in the presence of the exter-

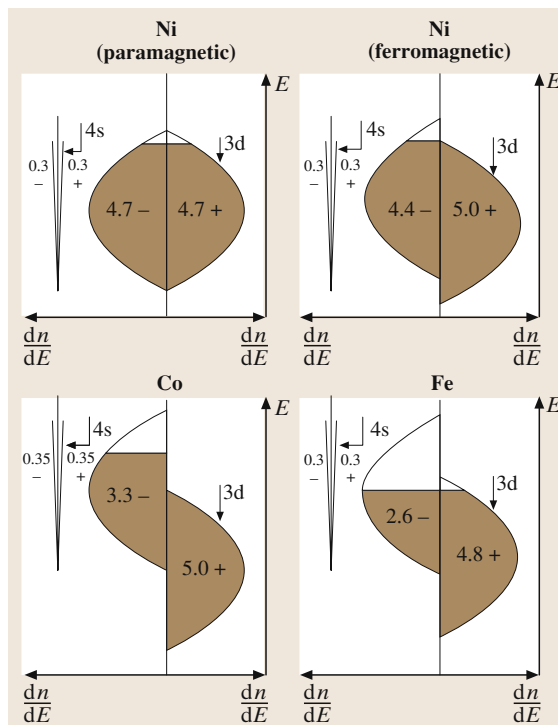


Fig. 51.15 Schematic of density of states curves for the ferromagnetic transition elements (Ni, Co and Fe). Both the 4s and 3d bands are shown. The magnetization of the elements is due to the difference between the spins in the spin-up and spin-down bands and the bands are full of electrons to the Fermi level (the shaded region). (After [51.30])

nal magnetic fields from the transitions in the recording medium.

The stack of films used in a spin valve has been discussed. The films used in Fig. 51.10, which we used to illustrate the GMR effect in a $\text{Co}_{90}\text{Fe}_{10}/\text{Ni}_{80}\text{Fe}_{20}$ free film, was: Ta (50 Å)/ $\text{Ni}_{80}\text{Fe}_{20}$ (60 Å)/ $\text{Co}_{90}\text{Fe}_{10}$ (55 Å)/Cu (variable)/ $\text{Co}_{90}\text{Fe}_{10}$ (55 Å)/FeMn (100 Å)/Ta (70 Å) [51.20]. In this case the antiferromagnetic film was FeMn (typically about 50%, 50%) and the Ta films are used to protect the rest of the stack.

Origin of GMR

The physical origin of the GMR effect lies in the differential scattering of electrons with spins in the same direction as the local magnetization (spin-up electrons) or opposite to this direction (spin-down electrons) at the interfaces and in the bulk part of the thin films comprising the GMR stack of films. The ferromagnetic transition elements (Fe, Co and Ni) have different 3d band structures, as shown in Fig. 51.15 [51.30] and discussed by [51.4]. This figure illustrates the energy of electrons in the two half-bands for spin-up and spin-down electrons versus the density of states $D(E)$ (number of states per unit energy per unit volume) in the 3d band for the ferromagnetic transition metals. The 4s bands are also shown. Electrons in the 4s bands make only a small contribution to the net spin but, because of the larger mobility (inversely proportional to the curvature of the bands), they are largely responsible for the current. The bands are filled to the Fermi energy at 0 K, and states above, but close to, the Fermi level are occupied at room temperature (illustrated by the shaded region in Fig. 51.15). The conductivity of a metal is given by

$$\sigma = \frac{1}{3} e^2 v_F^2 \tau D(E_F),$$

where e is the electron charge, v_F is the velocity of electrons at the Fermi energy, τ is the mean scattering time and $D(E_F)$ is the density of states at the Fermi energy [51.31]. If the mean scattering time (τ) is held constant then it is seen that the conductivity increases with the density of states at the Fermi energy; however, the scattering of the 4s electrons with a given spin is large when there are a large number of scattering centers and empty states with the same spin in the 3d band, i. e., the density of states at the Fermi energy $D(E_F)$ in the 3d band is large. In the ferromagnetic transition metals this is the largest effect and

$$\rho \propto D(E_F).$$

The scattering is due to spin–spin scattering, and spin–orbit scattering is believed to be small [51.32, 33]. The

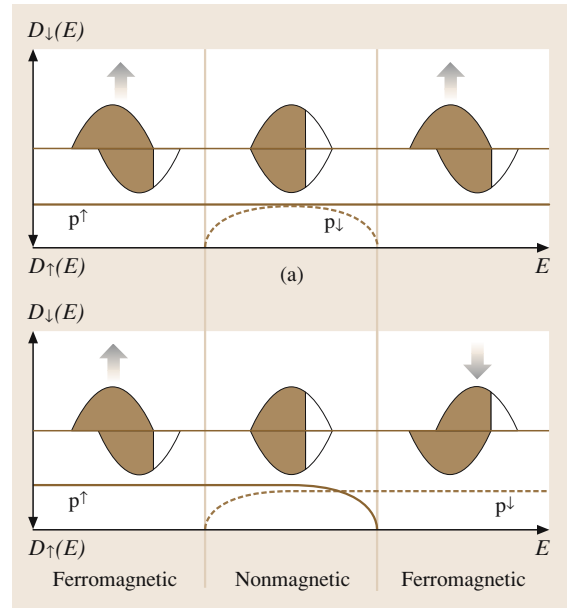


Fig. 51.16 Density of states curves for the spin valve structure in the ferromagnetic spin configuration (*top figure*) and the antiferromagnetic spin configuration (*bottom figure*). The resistivities of the layers are shown on the bottom part of the figures. (After [51.30])

magnetic moment per atom can be deduced as the difference in the number of electrons/atom with spin-up (the majority) versus those with spin-down. For example, the magnetic moment per atom for iron is $4.8 - 2.6 = 2.2$ Bohr magnetons per atom. From Fig. 51.15 we conclude that the differential scattering will be large in cobalt since for this element the difference in the density of states at the Fermi level is large [approximately 8–10 to 1 [51.30]]. A schematic of the density of states versus energy for the three films comprising the spin valve structure from the two extreme cases of the orientation of the ferromagnetic films (ferromagnetic configuration and antiferromagnetic configuration) is shown in Fig. 51.16 [51.30, 34]. All films have the same Fermi energy. Mathon described the GMR effect by discussing the differential scattering of *holes* in the 3d bands, but by concentrating on the scattering of electrons from the 4s to the 3d band we can make the same conclusions [51.32, 34]. It is assumed that the mean free path for the electrons is large compared with the thickness of any layer in the superlattice. Consider the ferromagnetic configuration case (top of Fig. 51.16): here the majority electrons have spin-up and the scattering of 4s electrons is low in both ferromagnetic films since both have a low

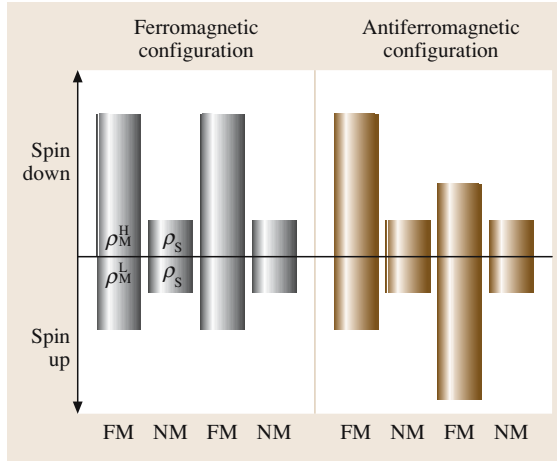


Fig. 51.17 Distributions of local resistivities for the spin-up and spin-down electrons in the ferromagnetic and antiferromagnetic spin configurations. (After [51.30])

density of states at the Fermi level. The nonmagnetic film has a larger density of states and will scatter electrons regardless of the orientation of the ferromagnetic films, which is why this film must be thin. Now consider the antiferromagnetic configuration case in Fig. 51.16: from the density of states at the Fermi level in the three films it is seen that the scattering is large in the right-hand ferromagnetic film for the electrons leaving the left-hand ferromagnetic film with spin-up (top density-of-states curve), resulting in large resistivity for this class of electrons. Similarly, the scattering of spin-down electrons is large for the left-hand ferromagnetic film. The

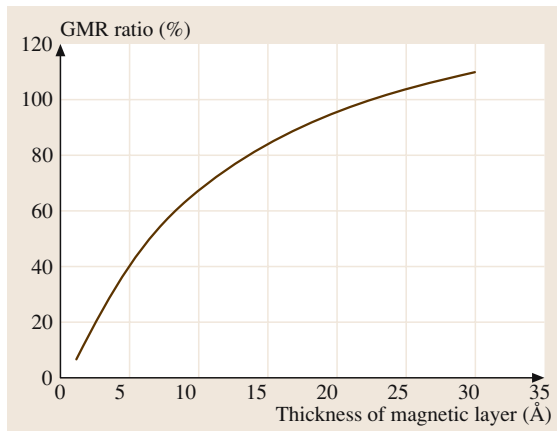


Fig. 51.18 $\Delta R/R$ versus the thickness of the magnetic layer (M) for a nonmagnetic layer thickness (N) of 10 Å for a Co/Cu spin valve. (After [51.30])

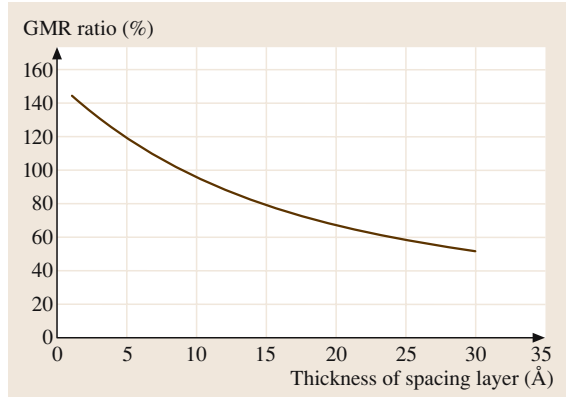


Fig. 51.19 $\Delta R/R$ versus the thickness of the nonmagnetic layer (N) for a magnetic layer thickness (M) of 20 Å for a Co/Cu spin valve. (After [51.30])

resistivity for the spin-up and spin-down electrons for the three films [ferromagnetic (FM), nonmagnetic (NM) and ferromagnetic (FM)] in the ferromagnetic and antiferromagnetic orientation cases is shown as a schematic in Fig. 51.17 [51.30] where it is shown that the ferromagnetic orientation of the ferromagnetic films has a (relatively) low resistivity path for the spin-up electrons while the resistivity for the antiferromagnetic orientation of the ferromagnetic films is high for both spin-up and spin-down electrons.

The change in resistance of the spin valve structure can be analyzed in terms of the different resistivities for the ferromagnetic and antiferromagnetic orientation of the two ferromagnetic films shown in Fig. 51.16 [51.30]

$$\frac{\Delta R_{\max}}{R_0} = (\alpha - \beta)^2 / [4 (\alpha + t_{\text{nonmagnetic}}/t_{\text{magnetic}}) \times (\beta + t_{\text{nonmagnetic}}/t_{\text{magnetic}})] , \quad (51.12)$$

where

$$\alpha = \frac{\rho_M^H}{\rho_s}, \quad \beta = \frac{\rho_M^L}{\rho_s},$$

where $t_{\text{nonmagnetic}}$ is the thickness of the nonmagnetic film and t_{magnetic} is the thickness of the magnetic film.

Figure 51.18 shows the variation of $\Delta R_{\max}/R_0$ for $t_{\text{nonmagnetic}} = 10$ Å versus t_{magnetic} for assumed values of $\alpha = 8.0$ and $\beta = 1.0$, which is appropriate for a Co/Cu superlattice [51.30]. For thicknesses of the magnetic films beyond approximately 10 Å, the change in $\Delta R/R$ decreases. Figure 51.19 shows the variation of $\Delta R/R$ for $t_{\text{magnetic}} = 20$ Å versus $t_{\text{nonmagnetic}}$. The decrease in

$\Delta R/R$ is rapid with increasing thickness of the nonmagnetic layer. This combination of films is not used in spin valves since the cobalt film is not magnetically soft; the usual choice is Permalloy with a variety of nonmagnetic transition metals and the resulting $\Delta R/R$ is lower.

51.1.4 Longitudinal Recording Media

Magnetic recording media used in modern disk drives are all sputter-deposited thin films of cobalt alloys for high coercivity and media flux ($M_r \times \delta$). Coercivities of 4000–5000 Oe are used (2006) and are limited by the ability of the write head to record on the media. The alloys that have evolved are four- or five-element alloys of CoPtCrX , where X is Ta or B or both. A typical atomic concentration of the four-element media is $\text{Co}_{64}\text{Cr}_{24}\text{Pt}_8\text{B}_4$. It has been found necessary to form a microstructure in the magnetic film that consists of a phase of small magnetic grains isolated by a nonmagnetic phase. The isolation of the magnetic grains results in less exchange coupling between grains and reduced noise – as we will discuss. Chemical segregation of $\text{Co}_{100-x}\text{Cr}_x$ alloys forms a nonmagnetic phase (Cr-rich) and a magnetic phase (Co-rich) and the nonmagnetic phase forms at the grain boundaries. The microstructure is controlled by undercoats that promote grain growth with the axes of the grains in the plane of the film. Platinum results in higher crystalline anisotropy and boron promotes the formation of an amorphous alloy of CoCrB , which grows at the grain boundaries, further

reducing the exchange coupling. Figure 51.20 shows the microstructure for a magnetic recording surface, illustrating the separation of the grains by the amorphous CoCr phase and, in the lower image, the c -axis stacking of the hcp grains [51.35]. In five-element alloys [CoCrPtBTa], Ta assists the segregation of Cr at the grain boundaries [51.36]. Some recording media use as many as three magnetic layers with the saturation magnetization increasing towards the read head, so that the average saturation magnetization is reduced while the magnetization nearest to the read head is increased. The layer of films comprising a recording layer at the areal densities for a 130-Gb/in² demonstration were: NiP-plated aluminium substrate with mechanical texturing in the data-recording zone/Cr seed layer/a CrMoX (X = grain-boundary segregation-enhancement element) underlayer/a CoCrT intermediate layer/two ferromagnetic layers (one with high Cr/low B concentration and the second with low Cr/high B concentration) and a double-layer amorphous carbon overcoat [51.37]. The substrate used for the samples in Fig. 51.20 was glass. Glass substrates have several advantages over aluminium with electroless Ni(P) (for hardness) substrates, including increased stiffness, which leads to reduced flutter, increased hardness and reduced susceptibility to corrosion. However, aluminium substrates continue to be used in nonportable applications because of reduced cost. The amorphous carbon is required to protect the magnetic film and to give a mechanically rigid surface on which the slider can fly at less than 25 nm.

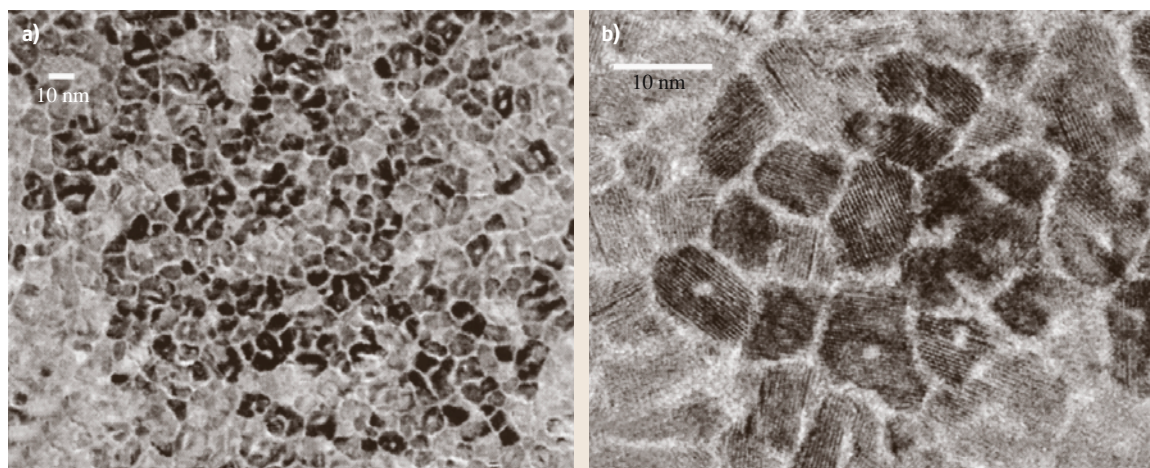


Fig. 51.20a,b Plan-view electron microscope images [the upper image (a) is lower resolution and the lower (b) is higher resolution] of the CoPtCrB magnetic alloy layer used in the 35-Gb/in² demonstration by IBM. The amorphous grain boundaries are seen as white and the c -axis stacking of the hcp grains can be seen within the grains in the higher-resolution image. (After [51.35].) The images were taken by Kai Tsang of IBM

Media Noise

In modern digital recording systems the dominant noise voltage which appears simultaneously with the desired read signal at the terminals of the read head is that from the recording media, assuming that the resistance from the read head is not excessive. Noise in the read channel is caused by random variations in the location of the recorded transitions. Transitions are not straight across the recorded track, as we assumed in our discussion of the transition parameter, but instead form zig-zag patterns in order to reduce the large demagnetizing field and demagnetizing energy associated with the longitudinal transition (discussed in [51.4]). Transitions also tend to follow the grain boundaries in the polycrystalline media. To achieve low media noise, it is required to have a small grain size so that the random nature of the grains is averaged over the width of the read head (the track width). An analysis of the signal-to-noise ratio, defined as the base-to-peak signal voltage amplitude (S_0) over the media noise voltage amplitude (N) has been published [51.24]

$$\left(\frac{S_0}{N}\right)^2 = 0.31 \left(\frac{B}{a}\right)^2 \left(\frac{PW_{50}}{B}\right) \frac{W_r}{W_c}, \tag{51.13}$$

- W_r : width of read head,
- PW_{50} : width of the read pulse at 50% of the peak amplitude,
- W_c : cross-track correlation width, which has been approximated as the grain or cluster size of the recording media. The grain size is a distribution and W_c is the mean of the distribution,
- a : transition parameter discussed above,
- B : length between bits.

For a reliable low error rate the value of $20\log S_0/N$ is required to be above 26–27 dB and this has required an increasingly smaller grain size as the width of the read head has decreased, resulting in higher track densities. For a 10-Gb/in² system the value of W_r was 0.505 μm and the required mean of the grain size was 12 nm [51.38]. The ratio B/a has been ≈ 3 to avoid severe distortion of the transition across the track, referred to as percolation (discussed in [51.4]) and the ratio PW_{50}/B has been in the region 2.5–2.8 for the type of detection channels used in present disk drives [partial response maximum likelihood (PRML)]. Therefore, to maintain constant S_0/N the ratio of the track width to the grain size must be kept constant and, with the decreased track width, the grain size must decrease in proportion.

Increased areal density has required decreased mean grain size but, equally important, tighter distributions

Table 51.3 Comparison of 20- and 80-GB/platter disk characteristics (after [51.17])

	20 GB/ 90 mm disk	80 GB/ 90 mm disk
H_c (Oe)	2700	4200
$M_r\delta$ (memu/cm ²)	0.46	0.32
Average grain size (nm)	10.8	9.1
Distribution sigma	0.41	0.24
Amplitude decay rate (%/decade)	0.7	0.4

of grain size. The distributions are closely matched by log-normal distributions of the form

$$y = \frac{1}{\sigma_Z \sqrt{2\pi}} \exp\left(-\frac{(Z - \bar{Z})^2}{2\sigma_Z^2}\right), \tag{51.14}$$

where Z is the log₁₀ value of the grain diameter, \bar{Z} is the mean value of Z and σ_Z is the standard deviation of Z . Grain size distributions are measured on foils by tunneling electron microscopy (TEM). Figure 51.20 showed the microstructure of a typical recording surface, emphasizing the segregation of the grains by the nonmagnetic constituents of the alloy. The distribution of the grain sizes in recording media for 20 GB/platter (counting both sides of the platter) and the higher density 80 GB/platter are shown in Table 51.3 [51.17]. Both the mean and standard deviation for the 80 GB/platter are reduced from those of the 20 GB/platter media. We will discuss the decay rates shown in Table 51.3 in the next section. The trend in the mean grain size and grains

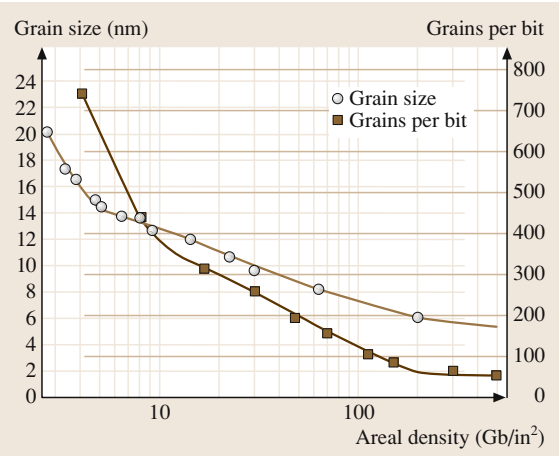


Fig. 51.21 Grain size and grains/bit for sputtered cobalt alloys as a function of areal density in Gb/in². (After [51.17])

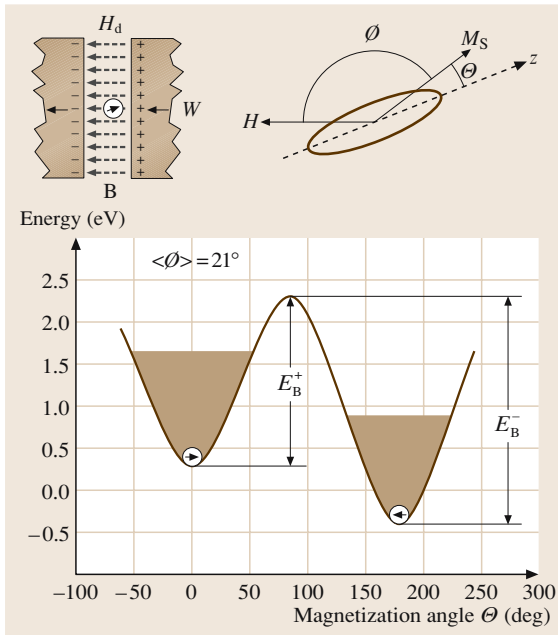


Fig. 51.22 Calculated energy of an isolated single-domain particle at the center of an 80-nm-long dibit transition. The properties of the particle are listed in Table 51.4. The ratio of the barrier height to thermal energy is 43 (see text) and the height $43 k_B T$ is shown shaded. There is margin between the top of the shaded region and E_B^+ and the particle is stable for the time t_x of 10 y. (After [51.39])

per bit for recording films versus areal density is shown in Fig. 51.21 [51.17]. A serious problem with this trend is the fact that the ratio of magnetic to nonmagnetic volume in the film is decreasing faster than the average grain size, which results in the number of grains per bit decreasing faster than the average grain size. The price to be paid for the increasingly small grain size is the stability of the recorded data over time as we discuss in the

next section. To investigate how far the areal density can be extended with longitudinal recording use has been made of the signal-to-noise ratio calculation to see if an areal density of 200 Gb/in² is feasible. To ensure that the signal-to-noise ratio does not degrade when extending the areal density from 60 to 200 Gb/in², it is required that: (a) the transition parameter be reduced from 12.8 to 6.2 nm, to achieve which the coercive field (H_c) must be increased from 4000 to 5000 Oe, and (b) the product of the magnetization and thickness ($M_r \delta$) be reduced from 0.32 to 0.2 memu/cm³. In addition, the head-medium spacing must be reduced from 30 to 15 nm [51.17].

Superparamagnetic Effect

To maintain the growth in areal density as shown in Fig. 51.1, it is necessary to reduce the physical dimensions of the recording-system components — and, to maintain the signal-to-noise ratio at levels required for satisfactory error rates, this includes reducing the mean grain size of the recording media. However, at some grain size the thermally driven fluctuations of the orientation of the magnetization in the isolated grains result in increased probability of the magnetization being switched from the desired orientation, as initially recorded, to the opposite direction. This instability of the magnetization is the *superparamagnetic effect*. The magnetic energy in an isolated grain is $K_u V$, where K_u is the uniaxial anisotropy consisting of crystalline, magnetostriction and shape contributions and V is the volume of the particle or grain with diameter D [$\pi(D/2)^2 \delta$]. Thermal energy supplied from the environment to the particle is $k_B T$, where k_B is the Boltzmann constant (0.8619×10^{-4} eV/K). An analysis of the superparamagnetic effect using an isolated small single-domain asymmetrical particle is summarized in Fig. 51.22 [51.39]. Here, the magnetic particle is assumed to be in the center of two transitions and the energy contour results from an analysis using only

Table 51.4 Energy of single-domain particle in magnetic field of two transitions (after [51.39])

Parameter	Value
B (dibit length)	80 nm
K_u (particle anisotropy energy, consisting of shape and crystalline anisotropy contributions)	1.5×10^6 erg/cm ³ (1.5×10^5 J/m ³)
$M_r \delta$	0.39 memu/cm ² (3.9 mA)
M_S	375 emu/cm ³ (375×10^3 A/m)
D (grain or particle diameter)	15 nm
δ (media thickness)	13 nm
V (grain or particle volume)	$\pi \times (D/2)^2 \times \delta$
H_d (demagnetizing field between transitions)	780 Oe (62 kA/m)
$\langle \phi \rangle$ (mean angle of field with respect to grain or particle axis)	21°

shape and crystalline anisotropy (combined into K_u) and magnetostatic energy terms.

$$E(\theta, \phi) = K_u V \sin^2 \theta + H M_s \cos(\theta - \phi) \quad (51.15)$$

It is assumed in Fig. 51.22 that the angle ϕ has a mean value $\langle \phi \rangle = 21^\circ$. The energy contour shown in Fig. 51.22 results from the application of (51.15) with the parameters listed in Table 51.4. The energy required to switch the particle from the initially magnetized direction is shown as E_B^+ . In the absence of an external or demagnetizing field, the energy barrier is

$$E_B^+ = K_u V.$$

With a demagnetizing field (H_d) the switching field is reduced since the field is in the direction of the switched magnetization

$$E_B^+ = K_u V \left(1 - \frac{|H_d|}{H_0} \right)^{1/2} \quad (51.16)$$

With a random two-dimensional (2-D) system $H_0 = 2K_u/M_s$. The magnetization of the particle will switch when $H_0 = H_d$. When the magnetization of the particle is switched, the new energy barrier is E_B^- and is increased in depth since the demagnetizing field is now in the same direction as the magnetization. Using the set of parameters listed in Table 51.4, the depth of E_B^+ [with a demagnetizing field of 780 Oe (62 kA/m)] is approximately 1.9 eV.

The time constant for the thermally driven switching of the magnetic particle in the presence of a magnetic field, which in the case under consideration is the demagnetizing field acting on the magnetic particle, can be found using an Arrhenius–Néel model

$$\tau = \tau_0 \exp \left(\frac{-E_B^+(H)}{k_B T} \right),$$

where $\tau_0 = 1/f_0$ and f_0 is an *attempt frequency*, usually taken as $\approx 10^9$ Hz [51.39]. The energy barrier depends on the thermal energy and the fraction x ($0 < x < 1$) of the retained magnetization and the storage time t_x

$$E_B^+(H) = \ln \left(\frac{t_x f_0}{|\ln x|} \right) k_B T, \quad (51.17)$$

$$t_x = |\ln x| (\tau_0) \exp \left(\frac{E_B^+(H)}{k_B T} \right).$$

As an example, if the required storage time is $\approx 10^8$ s (10y) and the percentage loss of data amplitude is 5%, then $x = 0.95$ and $|\ln 0.95| = 0.05$ and

$$\frac{E_B(H)}{k_B T} \cong 43.$$

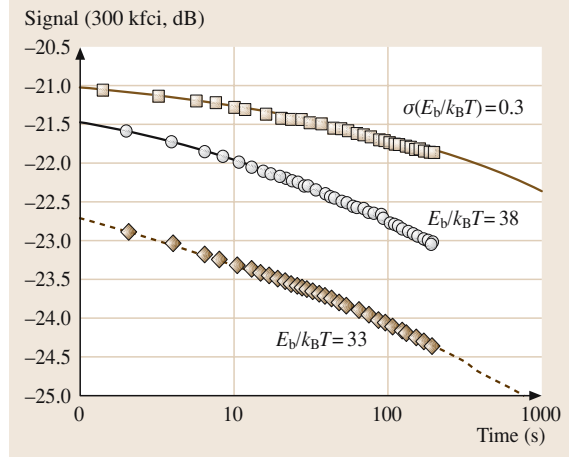


Fig. 51.23 Signal loss versus time at room temperature and 300 k flux changes per inch for three magnetic recording surfaces: The first two (from the top) are SFM and AFC media with layer thicknesses of: [12 nm/0.7 nm/4 nm] (open square), [12 nm/1.4 nm/4 nm] (open circle). The lower curve is for a 12-nm single layer (solid diamond). (After [51.40])

The thermal energy $k_B T$ at 350 K (the assumed operating temperature of a typical disk drive) is 0.03 eV, $43 \times k_B T \approx 1.25$ eV and the shaded region in Fig. 51.22 has this energy level. Since there is some margin between the thermally driven contribution to switching and the top of the barrier, the particle is stable (over the 10-year period). However, for smaller volumes (V) the value of E_B^+ decreases and the particle could become unstable. For example, if the thickness of the recording media is reduced from that assumed in Fig. 51.22 (13 nm) to 6 nm, the value of E_B^+ is reduced to ≈ 0.9 eV and the thermal energy would easily switch the particle.

A dramatic loss of data over time is observed with recording media with media flux ($M_r \times \delta$) less than about 0.25 memu/cm². Figure 51.23 shows amplitude loss of recorded data for three different media at a linear density of 300 k flux changes per inch at room temperature [51.40]. The three media are: a single-layer of CoCrPtB (lowest curve with a thickness of 12 nm); and two recording surfaces we will discuss in the next section. The signal decay rate R is defined as

$$R(H_d, T) = 100 \frac{\frac{A(t)}{A(t_0)} - 1}{\log(t/t_0)}, \quad (51.18)$$

where the signal amplitude of a square-wave pattern is $A(t)$. With $t_0 = 1$ s, the decay rate per decade for the media shown as the lower curve in Fig. 51.23 is

found to be -7.19% . The decay rate for the other media will be discussed in the next section. The decay rates for the two recording films shown in Table 51.3 are $-0.7\%/decade$ and $-0.4\%/decade$ for the 20 GB/platter and 80 GB/platter media, respectively. In a conventional disk recording system this superparamagnetic effect was judged to limit the areal density to somewhere in the range $36\text{--}100\text{ Gb/in}^2$ (reviewed in [51.39]). The superparamagnetic effect and the requirement for low media noise result in recording-system designers being caught in two inconsistent requirements: first, a small grain size for the recording media is required for adequate signal-to-noise ratio, so that the number of grains per bit remains constant (51.13), and second, a large grain size is required for small data loss over time because of the superparamagnetic effect [(51.16) and (51.17)]. The requirement to reduce the thickness of the recording media can be deduced from the need for low values of $M_r\delta$ for short transition lengths (51.5). The approach of increasing the anisotropy of the grains to increase $K_u V$ to resolving the superparamagnetic effect is limited by the ability of write heads to record on the media with the resulting increased media coercivity. The present values of areal density are approaching this limit. However, as can be seen from Table 51.3, the decay rates observed with the smaller grain size–higher density media is actually smaller than that for the larger grain size–lower density media. A possible explanation is that the $K_u V$ values for the smaller grains are higher than for the larger grains [51.17]. In the next section we discuss a novel approach that has extended the progression in areal density.

Antiferromagnetically Coupled (AFC) or Synthetic Ferrimagnetic (SFM) Media

The invention of a new form of recording media which has decreased the impact of the superparamagnetic effect was accomplished simultaneously by teams of researchers at the Fujitsu and IBM Almaden Research Laboratories [51.41, 42]. The new form of recording media consists of two ferromagnetic layers coupled by a thin layer of ruthenium [antiferromagnetically coupled (AFC) or synthetic ferromagnetic media (SFM)], which results in an antiferromagnetic coupling of the two films, similar to that used in the spin valve for a synthetic antiferromagnet. In order for the coupling between the two ferromagnetic layers to be antiferromagnetic, the thickness of the ruthenium layer, referred to by IBM as “pixie dust”, is required to be approximately 0.6 nm [51.42]. A schematic of the antiferromagnetic coupled media and the resulting hysteresis curve is

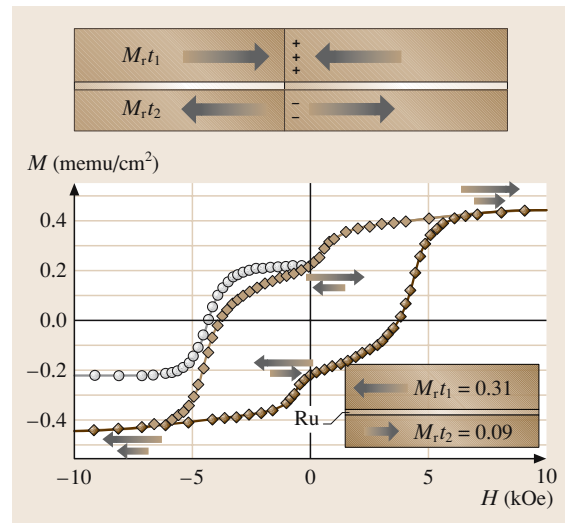


Fig. 51.24 (top) Schematic of the antiferromagnetically coupled (SFM or AFC) media showing the two ferromagnetic layers coupled by a thin layer of ruthenium (“pixie dust”) and with a single magnetic transition for which the magnetizations in the two layers are oppositely directed. (bottom) Magnetic hysteresis curve for the SFM or AFC media with remanent thickness products for the two ferromagnetic layers shown in the inset and measured at 350 K. The filled triangles are the major hysteresis curve and the filled circles are the remanence hysteresis curve. The arrows show the directions of the magnetization in the two films. (After [51.40])

shown as Fig. 51.24 [51.40]. The sharp reduction in the remanent magnetization at $\approx 80\text{ Oe}$ is due to the switching of the thinner lower layer ($M_r t_2$) and the lowered value of the remanence is advantageous for reduced transition noise and transition length. The curve with solid circles is a remanence hysteresis curve measured by subjecting the sample to a negative reversing field and then increasing the field in a positive direction to zero and plotting the remanent magnetization. This measurement is an attempt to simulate what happens when a recording medium leaves the vicinity of the recording head. By analyzing the read pulse from the transition in the upper layer combined with the read pulse from the transition with reversed sign from the lower layer (the layers are antiferromagnetically coupled), it can be shown that no degradation of the resolution (PW_{50}) of the system results from having two as opposed to one recording layer [51.42]. The advantage of the AFC or SFM configuration is that the medium is in fact thicker than would be required for the given recording density using a single

layer and hence is less susceptible to the superparamagnetic effect (51.17). The effective volume is bounded by

$$K_uV_1 < K_uV_{\text{eff}} < K_uV_1 + K_uV_2$$

and with tight coupling and alignment of the axes of the grains the upper limit is approached. The effective remanent magnetization \times thickness product is close to

$$M_{\text{r}}t_1 - M_{\text{r}}t_2$$

and this reduces the pulse width over that a single-layer recording medium (51.10).

Because of the antiferromagnetic coupling, an exchange magnetic field $H_{\text{ex},2}$ acts on the lower magnetic film

$$H_{\text{ex},2} = \frac{J_{\text{ex},2}}{M_2t_2}$$

and a typical value for the antiferromagnetic exchange constant $J_{\text{ex},2}$ is 0.06 erg/cm³ (6 mJ/m³) [51.44]. For the magnetization of the lower layer to assume the opposite direction to that of the top layer before the field H is reduced to zero, it is required that $H_{\text{ex},2} > H_{\text{c}2}$. The antiferromagnetic exchange constant can be increased by adding a thin layers of cobalt or cobalt alloys known as E layers to both sides of the ruthenium layer [51.43]. The increase in the antiferromagnetic exchange constant is rapid with thickness of the E layers and we show in Table 51.5 a listing of J_{ex} for several E layers [51.43]. The largest increase in J_{ex} is with cobalt, but this material suffers from significant degradation on overwrite and poor signal-to-noise ratio compared with cobalt alloys.

To illustrate that the AFM or SFM media configuration actually results in reduced amplitude loss over time, we show in Fig. 51.23 the amplitude loss with two SFM media: 12/0.7/4 (open squares) and 12/1.4/4 (open circles), where the numbers are the thickness of the three layers in nm and the value of $E_{\text{B}}/k_{\text{B}}T$ is 38 [51.40]. The signal decay rate is -4.14% per decade for the top curve (open squares) and -6.45% per decade for the middle curve (open circles) [51.40]. The coupling between the

two magnetic films is antiferromagnetic for the top curve and ferromagnetic for the middle curve and the lower curve is a single magnetic layer. The AFM or SFM media clearly reduces the decay rate. The improvement is even more significant with a larger value for $E_{\text{B}}/k_{\text{B}}T$ and values of R less than 1%/decade with $M_{\text{r}}t$ values lower than 0.2 memu/cm² have been reported by IBM researchers using static magnetization measurements with a reversing magnetic field to mimic the demagnetizing field [51.42]. AFM media are being used in IBM portable disk-drive applications [51.45]. Areal densities larger than 100 Gb/in² have been reported with longitudinal recording using AFM media [51.46]. The magnetic recording layers were alloys of CoCrPtB with an average grain size of 9.4 nm. The write head was trimmed on the air-bearing surface to the final track width using a focused ion beam and the track density was 149 ktracks/in and the bit density was 720 kbits/in [51.46].

The data shown in Table 51.3 also shows that it is possible to achieve low decay rates with single recording layers [51.17]. Analysis comparing SAF and single-layer media has shown that the single-layer media is preferred unless the media flux drops below about 0.25 memu/cm² with H_{c} values constrained below 4300 Oe. If recording heads are able to record with adequate overwrite on media with coercivity greater 4300 Oe, then the single-layer media is preferred [51.47].

Dynamic Coercivity

Another manifestation of thermal effects on magnetic properties is a change in coercivity with switching time. With long periods of time the probability that a particle will switch is increased by thermal excitation and with short periods this probability is decreased – leading to a change in the coercivity of the particle. An experimental study of this effect was done on a test stand in which a stationary write head is driven by a current pulse of variable duration over a previously magnetically saturated recording medium. The current pulse records two transitions (a dibt). The width of each transition is essentially the gap of the write head and the length of

Table 51.5 Comparison of materials for E layers: J_{ex} values and degradation in overwrite and signal to media noise (after [51.43])

Material for E layers (all 1 nm thick)	Co	Co–Cr	Co–Cr–X
J_{ex} (erg/cm ²)	0.73	0.59	0.24
Degradation in overwrite (dB) with respect to SFM media without E layers (negative values indicate improved overwrite)	10	≈ 3	≈ -6
Degradation in S/N _m (dB) with respect to SFM media without E layers	≈ 4.5	≈ 1.6	≈ 1.5

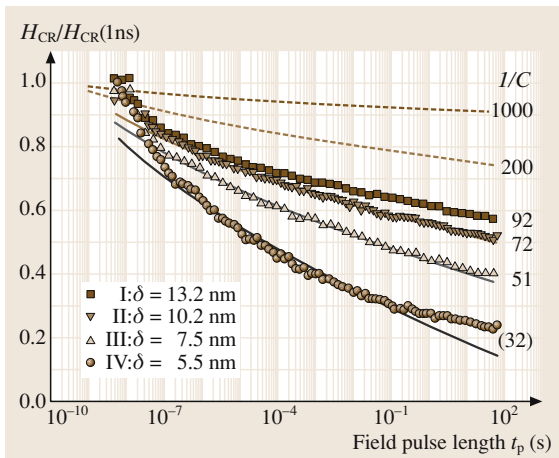


Fig. 51.25 Dynamic coercivity $H_{CR}(t)$ for a series of $\text{Co}_{64}\text{Pt}_{14}\text{Cr}_{22}$ films with variable thickness. (After [51.39])

the reversed magnetization region varies with the pulse width (t_p). The dibit is read with a magnetoresistive sensor moved across the recorded transitions resulting in a dibit pulse. By measuring the amplitude of the dibit signal as a function of the pulse width and varying the magnitude of the reverse field and measuring $H_{\text{head}}(I)$ it is possible to deduce the remanent coercivity (H_{CR}) as a function of t_p . The resulting curves of $H_{CR}/H_{CR}(1 \text{ ns})$ are shown in Fig. 51.25 [51.39]. The experimental slopes C increase as the thickness decreases. The inverse slope parameter is given by

$$\frac{1}{C} \cong \frac{K_u V}{k_B T},$$

which is the superparamagnetic stability parameter discussed above and it is observed that, even with a stability factor larger than 51, for which the superparamagnetic region is avoided, the coercivity increases significantly for short switching times. This is the actual coercivity that must be exceeded by the write-head magnetic fields to switch a recording film.

51.1.5 Perpendicular Recording

The technology that is now starting to supplant longitudinal recording is perpendicular recording. Perpendicular recording offers a way out of the conundrum caused by the need for thinner recording media for improvements in recording density and the need for larger particles to reduce the impact of superparamagnetism. In perpendicular recording the magnetization in the recording media is held perpendicular to the surface

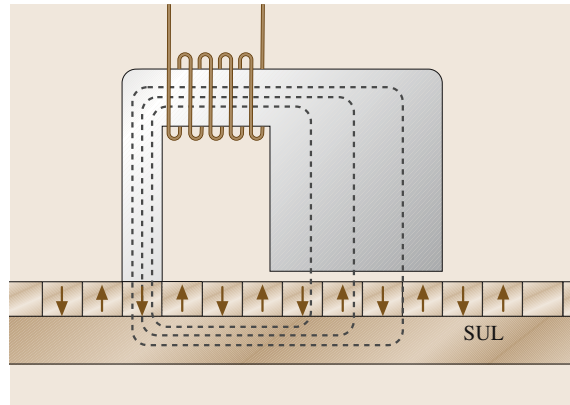


Fig. 51.26 Schematic of a perpendicular recording system. (After [51.4])

of the recording media by a perpendicular anisotropy large enough to overcome the large demagnetizing field with this orientation of the magnetization. A schematic of a perpendicular recording system including a probe write element and a magnetically soft underlayer (SUL) is shown as Fig. 51.26. It is also possible to have a perpendicular recording system with a ring write head and a single-layer media, but this approach has not been developed. In contrast to longitudinal magnetic recording the SUL is required for a low-reluctance path for the magnetic flux generated by the coil on the write element. One advantage of perpendicular recording is that the magnetic field for recording is larger than that for longitudinal recording since the magnetic field from the write pole is in the gap of the write head, in contrast to longitudinal recording where the field leaks from the write gap. A schematic of the total geometry for longitudinal and perpendicular recording is shown in Fig. 51.27 [51.1]. Reading for perpendicular recording is accomplished with a GMR sensor as with longitudinal recording. The major difference in the write head is that it uses a single pole to generate the perpendicular magnetic field with a wider (to reduce the reluctance) return pole. The write head for perpendicular recording is imaged in the SUL and this results in an increase in the strength of the magnetic field by approximately a factor of two. In addition, due to imaging effects, the effective number of turns on the write head is doubled for perpendicular recording. It is important for perpendicular recording that the magnetization of the magnetic material for the SUL be larger than that for the pole of the write head [51.48]. This is the case to prevent the SUL from saturating before the pole tip saturates, resulting in a decrease in the gradient of the head field, which would result in an

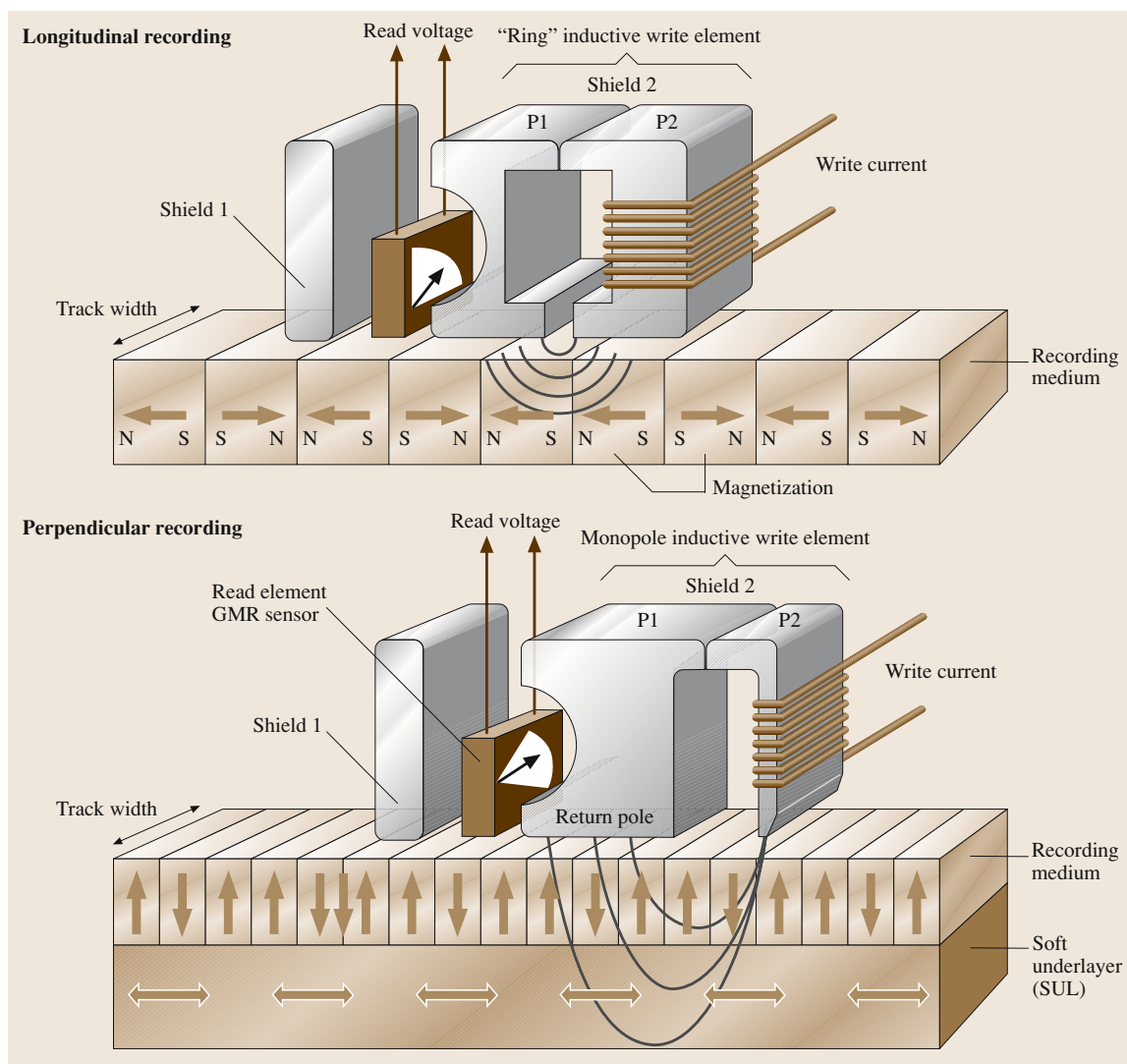


Fig. 51.27 Schematic of longitudinal and perpendicular recording systems. (After [51.1])

increased transition length (to be discussed). A large magnetic field and gradient of magnetic field from the write head can record media with a larger coercive field than for longitudinal recording, which results in a reduced transition length. Perpendicular-recording media can be made thicker than for longitudinal media for the same areal density. The media thermal stability factor ($\frac{K_u V}{k_B T}$) is increased both by an increase in K_u by virtue of the increased coercive field but also by the increase in the volume (V) with the thicker media. Reading of the magnetic transitions with perpendicular recording is done with a spin valve head as with longitudinal

recording – the difference being that with perpendicular recording the response is a step instead of a pulse (to be discussed).

A significant recent development for perpendicular write heads is the addition of a trailing-edge shield to the head that results in an improvement in the concentration of the writing flux and an increase in the gradient of the head field (dH_y/dx), which reduces the transition length [51.49,50]. A schematic of a write head with a trailing-edge shield is shown in Fig. 51.28 [51.51]. The shield connects to the larger return structure for the head shown in Fig. 51.27. The particular geometry

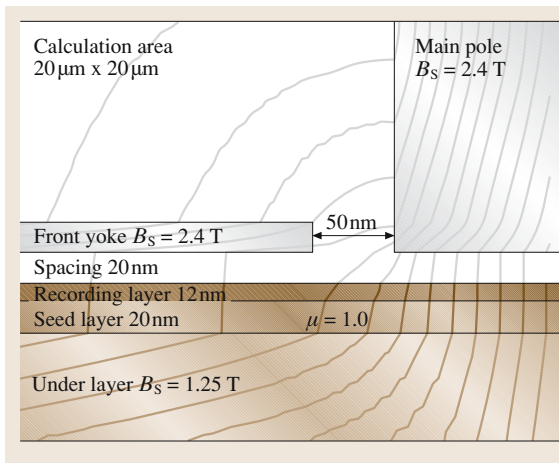


Fig. 51.28 Schematic of a perpendicular recording head including the main pole and the front yoke or trailing-edge shield. (After [51.51])

and saturation magnetization of the main pole and the *front yoke* or trailing-edge shield were used in a two-dimensional finite element model to calculate the head fields and shape of the recorded transition [51.51]. Results of the simulation of transition shape are shown in Fig. 51.29, where it is seen that the recorded transition is considerably sharper with the head with the front

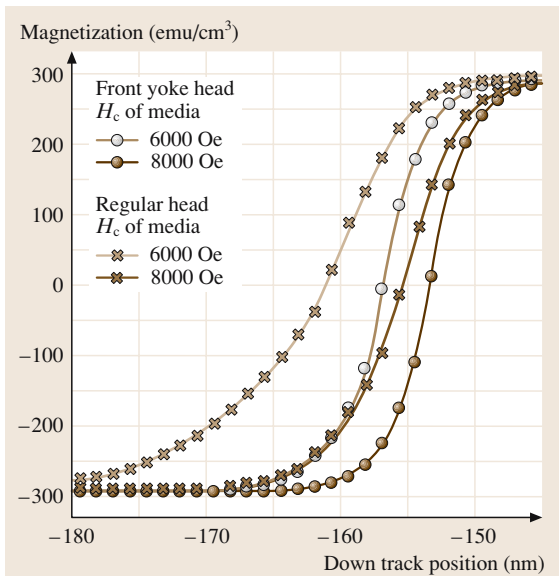


Fig. 51.29 Magnetic transitions in perpendicular recording with and without the front yoke or trailing-edge shield. (After [51.51])

yoke compared with the head without the shield, particularly with the higher coercivity media ($H_c = 8000$ Oe). The perpendicular magnetic field has the amplitude reduced to one-half at a distance of $0.8d$ with the shield and $1.3d$ without the shield (assuming that the spacing of the shield is spaced by a distance d from the main pole) [51.50]. The shield shunts some of the write flux, reducing the amplitude of the write field, so it is important to keep the height of the shield short. Another advantage of the trailing-edge shield is that a horizontal component of the field is added. This field reduces the field for switching since, if the angle of the field is θ , the switching field is

$$H_{\text{switch}} = \frac{H_k}{[(\cos \theta)^{2/3} + (\sin \theta)^{2/3}]^{3/2}},$$

where H_k is the anisotropy field of the perpendicular medium. The denominator of this equation is always larger than one and hence the switching field is reduced for angles θ larger than zero degrees. Note that this equation is the switching asteroid equation used to explain the operation of the MRAM (to be discussed).

A significant advantage of perpendicular as compared with longitudinal recording is the narrower off-track read profile as shown in Fig. 51.30 [51.50]. This improvement in read profile leads directly to increased track density in disk drives where there is significant skew of the head with respect to the track center line. Head skew results from the fact that present disk drives

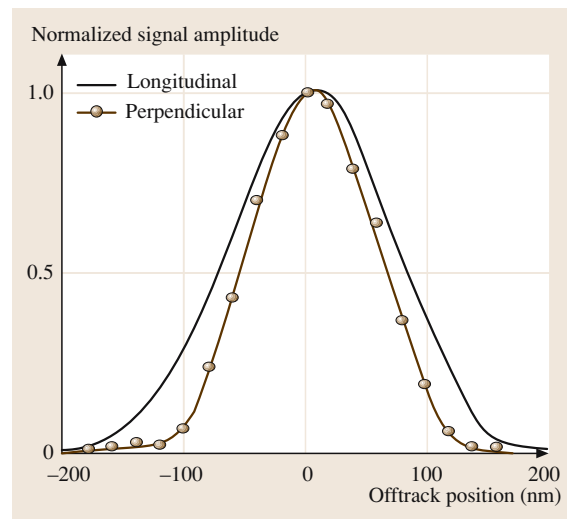


Fig. 51.30 Cross-track profile of the signal with perpendicular and longitudinal recording. (After [51.50])

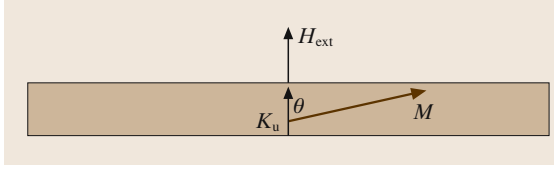


Fig. 51.31 Magnetization and perpendicular anisotropy for a perpendicular recording media. (After [51.4])

use rotary actuators and the angle of the written transition with respect to the track center line varies as the actuator is rotated across the disk, being largest at the inside and outside diameters. With head skew it is possible for the write element of the head to be placed over the guard band or adjacent track of a selected track, resulting in unwanted recording in these areas. This problem is reduced by using trapezoidal poles [51.52]. Shields can be added to the side of the head to reduce this erasure of adjacent track information; however, the processing of such shields is complex.

Media for Perpendicular Recording

The magnetization and anisotropy in a perpendicular recording film is shown as Fig. 51.31. The energy of the magnetization in the presence of the demagnetizing, anisotropy and magnetostatic energies (cgs units) is

$$E = \frac{1}{2} 4\pi M_s^2 \cos^2 \theta + K_u \sin^2 \theta - H_{\text{ext}} M_s \cos \theta, \quad (51.19)$$

where the demagnetizing factor for a thin film magnetized normal to the surface is 4π . The equilibrium angle is given by the solution to

$$\frac{dE}{d\theta} = \frac{4\pi}{2} (-2 \cos \theta \sin \theta) M_s^2 + 2K_u \sin \theta \cos \theta + H_{\text{ext}} M_s \sin \theta = 0. \quad (51.20)$$

For stability

$$\frac{d^2 E}{d\theta^2} = -4\pi (\cos^2 \theta - \sin^2 \theta) M_s^2 + 2K_u (\cos^2 \theta - \sin^2 \theta) + H_{\text{ext}} M_s \cos \theta > 0. \quad (51.21)$$

The solution is $\sin \theta = 0$ ($\theta = 0^\circ$ or 180°) and the stability condition for $H_{\text{ext}} = 0$ results in

$$K_u > 2\pi M_s^2. \quad (51.22)$$

Alloys under consideration for perpendicular recording include those listed in Table 51.6 [51.53] and Table 51.7. A review of work on perpendicular recording media was published [51.54]. It is observed from this table that these candidate materials satisfy the condition (51.22). The critical magnetic field below which the $\theta = 0^\circ$ solution is no longer stable is given by the solution to (51.21) with $H_{\text{ext}} \neq 0$, resulting in

$$H_{\text{e,crit}}(\theta = 0^\circ) = -\frac{(2K_u - 4\pi M_s^2)}{M_s},$$

$$H_{\text{e,crit}}(\theta = 180^\circ) = \frac{(2K_u - 4\pi M_s^2)}{M_s}. \quad (51.23)$$

The solution to (51.21) with $\sin \theta$ factored out is

$$-4\pi M_s^2 \cos \theta + 2K_u \cos \theta + H_{\text{ext}} M_s = 0,$$

$$H_{\text{ext}} = \cos \theta (4\pi M_s - 2K_u/M_s),$$

$$M_{\perp}/M_s = \cos \theta,$$

$$M_{\perp}/M_s = \frac{H_{\text{ext}}}{(4\pi M_s - 2K_u/M_s)}. \quad (51.24)$$

This solution applies when the magnetization curve is not saturated. $H_{\text{e,crit}}$ is not the same as the coercive field where $M(H_{\text{ext}}) = 0$. The coercive field is found from (51.20) by recognizing that the magnetic field $H_{\text{ext}} = -|H_{\text{e,crit}}| - H'$, where H' is measured from $-|H_{\text{e,crit}}|$. The coercive field is then

$$H_c = -2|H_{\text{e,crit}}|.$$

The result for a Co–Cr–Pt alloy (Table 51.7) with the following parameters: $K_u = 1.45 \times 10^6$ erg/cm³, $M_s = 280$ emu/cm³ is as follows

$$H_{\text{e,crit}}(\theta = 0^\circ) = -\frac{(2K_u - 4\pi M_s^2)}{M_s},$$

$$H_{\text{e,crit}}(\theta = 0^\circ) = -\frac{(2 \times 1.45 \times 10^6 - 4\pi 280^2)}{280},$$

$$H_{\text{e,crit}}(\theta = 0^\circ) = -6838 \text{ Oe},$$

$$H_{\text{e,crit}}(\theta = 180^\circ) = 6838 \text{ Oe}.$$

Table 51.6 Magnetic properties of selected materials for perpendicular recording (data from [51.53])

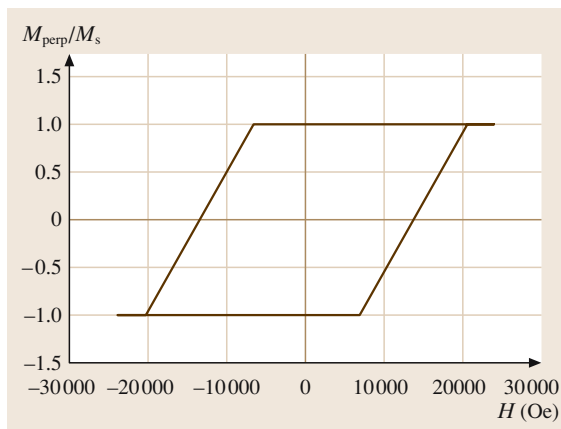
Material	H_k (kOe)	M_s (emu/cm ³)	K_u (erg/cm ³ × 10 ⁶)	$2\pi M_s^2$ (emu/cm ³) ²
Co–Cr	7	270	0.95	0.46
Co–Cr–Pt	10	280	1.45	0.49
Co/Pd	17	350	2.98	0.77
Fe/Pt	28	750	10.5	3.53

Table 51.7 Media for perpendicular recording

Media	H_c (kOe)	M_r (emu/cm ³)	α ($4\pi M_r/H_c$)	H_k (kOe)	Magnetic layer thick- ness (nm)	H_n (kOe)	Comments	Ref.
CoCrNbPt/Ti	2–5	250–450	1.25	6	50	–	Good SNR, poor thermal stability	[51.55]
CoCrPt/Ti/CoZr	2–4	500	2.09	13	25	–	Good SNR, poor thermal stability	[51.56]
Co _{70.3} Cr ₁₅ Pt _{14.7} /Ti/CoZrNb	5.5	460	1.05	–	–	–1.7	Good thermal stability, SNR = 25 dB	[51.57]
[Co/Pd] _n	2–10	200–600	0.84	15–30	10–50	–	Good thermal stability, poor SNR	[51.58]
[Co/PdSi] ₁₃	3.2–5.4	160	1.35	–	–	–	SNR improved	[51.59]
FePt	2–5	800	2.87	> 30	10–50	–	–	[51.60]
CoCrPtB/Ti–CoCrTa/laminated SUL	2.5–3.0	400	1.83	12	–	–	Good SNR, poor thermal stability	[51.61]
CoCrPtO/Ti–Ru/laminated SUL	3.5	520	1.86	–	–	–	Good thermal stability, poor SNR	[51.62]
CoPtCr–SiO ₂	> 5	430	1.08	18.6	10–30	–	Good thermal stability, good SNR	[51.63]
Coupled/granular/continuous CoCrPt/Co–Pt	4–5	400	1.11	–	–	–1 to –3	Good thermal stability, good SNR	[51.64]
Coupled/granular/continuous CoCr ₁₈ Pt ₁₂ /Co–Pt	4.02	395	1.23	–	–	–1.8	Good thermal stability. SNR = 31 dB	[51.65] [51.66] [51.67]

The magnetization curve for the second quadrant is:

$$M_{\perp}/M_s = \frac{H_{\text{ext}}}{(4\pi M_s - 2K_u/M_s)},$$

**Fig. 51.32** Hysteresis curve for Co–Cr–Pt perpendicular recording media using the analysis in the text

$$M_{\perp}/M_s = \frac{(H_{\text{ext}} + 2 \times |H_{e,\text{crit}}|)}{|H_{e,\text{crit}}|},$$

$$H_c = -2 \times |H_{e,\text{crit}}| = -13.67 \text{ kOe}.$$

The magnetization curve for the fourth quadrant is:

$$M_{\perp}/M_s = \frac{(H_{\text{ext}} - |H_{e,\text{crit}}|)}{|H_{e,\text{crit}}|} - 1.$$

The magnetization curve is shown in Fig. 51.32 and exhibits the desired large coercivity but the curve is sheared due to the large demagnetizing field.

In general, the magnetic hysteresis curve (M versus H_{ext}) for the recording medium is as illustrated in Fig. 51.33 [51.53]. Here the slope parameter (α) characterizes the slope of the magnetization curve at $H = -H_c$,

$$\alpha = 4\pi \left(\frac{dM}{dH} \right)_{H=-H_c}. \quad (51.25)$$

Finite values of α result from variations in the direction of the perpendicular magnetization, dispersion in the

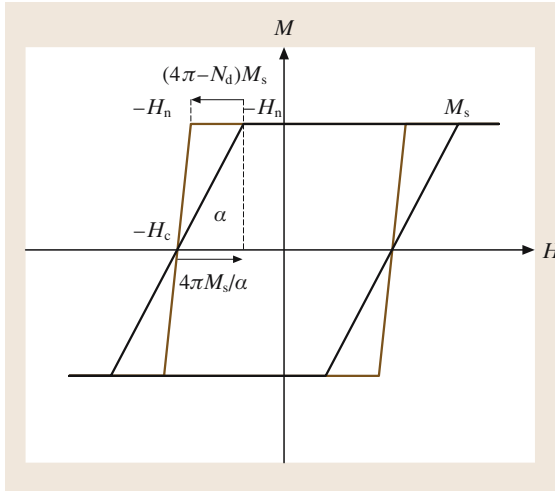


Fig. 51.33 Hysteresis curve with demagnetizing field ($4\pi M_s/\alpha$) and a reduced demagnetizing factor N_d for a perpendicular recording film with loop slope parameter α . (After [51.53])

perpendicular anisotropy and variations in the exchange coupling between grains and clusters in the media. The effective demagnetizing field is reduced from $4\pi M$ to $4\pi M/\alpha$. This reduction in demagnetizing field results from the formation of a rectangular checkerboard pattern of domains with high-density linear recording and high track densities [51.53]. The effective anisotropy energy, which assists in overcoming the thermal energy and the superparamagnetic limit, is given in terms of the nucleation field (H'_n), at which field the magnetization just starts to reverse, by

$$K_{\text{eff}} = H'_n M_s / 2.$$

For high areal density an estimate of the reduced demagnetizing factor is

$$N'_d = 0.5 \times 4\pi M_s$$

and the increase in the nucleation field H_n is

$$H'_n = H_n + \frac{4\pi M_s}{2}.$$

For high-density recording large values for α and H_c are required; however, the value of H_c is limited by the ability of the pole head to generate magnetic fields equal to or greater than approximately $1.5 \times H_c$, the condition that ensures adequate overwrite.

The transition parameter (a) for perpendicular recording can be estimated using the same write-slope

Table 51.8 Parameters for evaluating transition parameters for longitudinal and perpendicular recording

Parameter	a_L	a_P
d (nm)	26	26
t (nm)	—	50
u (nm)	—	0
S^*	0.95	0.98
M_r (emu/cm ³)	200	200
δ (nm)	14	—
Q	0.6	0.6
$H_{c,\text{max}}$ (Oe)	3500	8000

criteria discussed for longitudinal recording [51.68]. Here,

$$\frac{dM}{dx} = \frac{dM}{dH} \left(\frac{dH_{\text{head}}}{dx} + \frac{dH_{\text{demagnetizing}}}{dx} \right). \quad (51.26)$$

This equation is solved by assuming an arctangent transition shape and denoting the slope of the head field by $\frac{dH_{\text{head}}}{dx} = QH_c/s$, where s is the total spacing of the write pole tip to the top of the SUL, consisting of the physical spacing (d), the thickness of the recording media (δ) and the spacing between the SUL and the magnetic recording film (u). Using the demagnetizing field for a perpendicular transition [51.69]

$$a = a_1 + \sqrt{\left(a_1^2 + \frac{(1-S^*)\delta s}{\pi Q} \right)},$$

$$a_1 = \frac{4M_r s}{QH_c} + \frac{s(1-S^*)}{\pi Q} - \frac{\delta}{4}, \quad \alpha = \frac{4\pi M_r}{H_c(1-S^*)}. \quad (51.27)$$

A comparison between the transition parameters for longitudinal (a_L) and perpendicular (a_P) recording from (51.5) and (51.27) with the parameters listed in Table 51.8 is shown in Fig. 51.34 [(H_c) for longitudinal recording is the dark line and (H_c) for perpendicular recording is dashed]. It is seen that the transition parameter for perpendicular recording can be made smaller than that for longitudinal recording if the coercive field for perpendicular recording is larger than 6500 Oe. Note that the thickness of the recording media for perpendicular recording is larger than that for longitudinal recording. The magnetic squareness (M_r/M_s) for perpendicular recording media is larger than that for longitudinal recording.

Gao and Bertram have shown a significant advantage for perpendicular recording if the uniaxial anisotropy is not aligned perpendicular to the surface, but oriented at

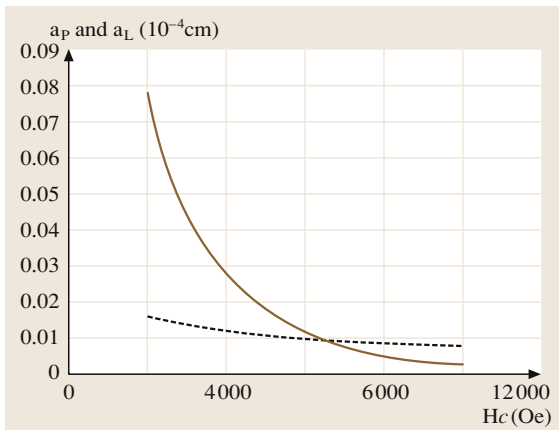


Fig. 51.34 Comparison of transition parameters versus the coercive field (H_c) for perpendicular (solid line) and longitudinal (dashed line) recording. (After [51.68])

45° to the perpendicular [51.70]. The geometry for this *tilted* perpendicular recording is shown as Fig. 51.35. The crystallites comprising the media are still aligned perpendicular but the crystalline anisotropy is assumed to be at 45°. No material system was disclosed that exhibits this tilted anisotropy. With tilted perpendicular recording the crystalline anisotropy field (H_k) can be made significantly larger than if the anisotropy is perpendicular. An example is 15 000 Oe for perpendicular recording and 28 000 Oe for tilted perpendicular recording [51.70]. Some of the advantages of tilted perpendicular recording are: reduced switching field and reduced average grain diameter, resulting in increased signal-to-noise ratio since the number of grains per bit is increased.

The requirements for media for ultra-high-density perpendicular recording have been analyzed [51.71]. Table 51.9 summarizes the results for media parameters for areal densities of 100–1000 Gb/in². Several key media parameters are the grains per bit (N) and the coercivity (H_c). An estimate of the signal-to-noise (SNR) ratio using the Poisson statistics of additive noise for the magnetic grains is

$$\text{SNR} = -10 \log_{10} (N) ,$$

where N is the number of magnetic grains per bit [51.71]. There are two approaches to realizing recording media with a large perpendicular anisotropy and that can approach the parameters in Table 51.9: granular media typified by CoCr alloys and *continuous* multilayer media typified by [Co–Pd]_n multilayers. CoCr alloys exhibit good SNR but insufficient H_c for ultra-high recording

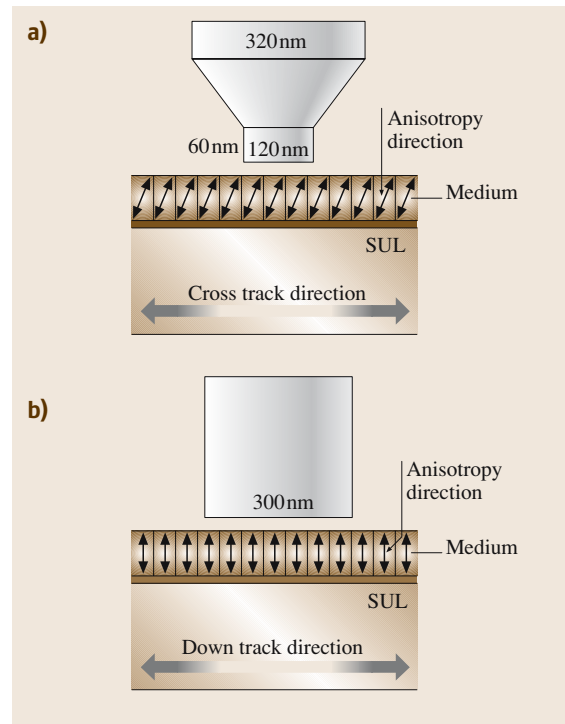


Fig. 51.35 Typical geometry of tilted perpendicular recording. The medium anisotropy directions and not the medium grains are tilted in the cross-track direction. (After [51.70])

densities, while the multilayers have not exhibited sufficient SNR. Recent results for multilayers of Co/PdSi where the Pd layer is doped with silicon have shown an improvement in the SNR [51.59]. The technology demonstration of 50 + Gb/in² (to be discussed) used a granular CoCr alloy. Table 51.7 is a listing of some metal alloy candidates for perpendicular recording. In general, Pt plays a significant role in determining the magnetic anisotropy and Cr, Ta, Nb and B reduce the exchange coupling, resulting in reduced noise. Another alternative for perpendicular recording media is oxide media, e.g., CoCrPtO or CoCrPt–SiO₂ [51.63, 72, 73]. A recent study has shown that the oxide media has improved coercivity and a negative nucleation field (H_n), reduced pulse width and reduced signal decay compared to CoCrPtB media [51.74]. An issue with the oxide media is the nonuniformity of the coercivity and oxygen content in the disk surface due to the high gas pressure and oxygen-gas reactive sputtering [51.74]. The last two listings are for a *coupled granular/continuous* (CGC) medium that combines the best properties of the continuous media (high H_c and good thermal stability) and gran-

Table 51.9 Requirements for perpendicular recording media for ultrahigh recording densities (after [51.71])

Areal density (Gb/in ²)	100	250	500	1000
Flying height <i>d</i> (nm)	8	7	6	5
Signal-to-noise ratio (RMS signal/RMS noise)	18	17	16	15
Grains per bit <i>N</i>	63	50	40	32
Average grain diameter <i>D</i> (nm)	9	7	5	4
<i>S</i> *	1	1	1	1
Media thickness <i>δ</i> (nm)	18	20	22	24
Perpendicular anisotropy constant <i>K</i> _u (×10 ⁶ erg/cm ³)	2	3	4	6
Saturation magnetization <i>M</i> _s (emu/cm ³)	400	550	700	875
Coercivity <i>H</i> _c (Oe)	8500	11000	12500	14300

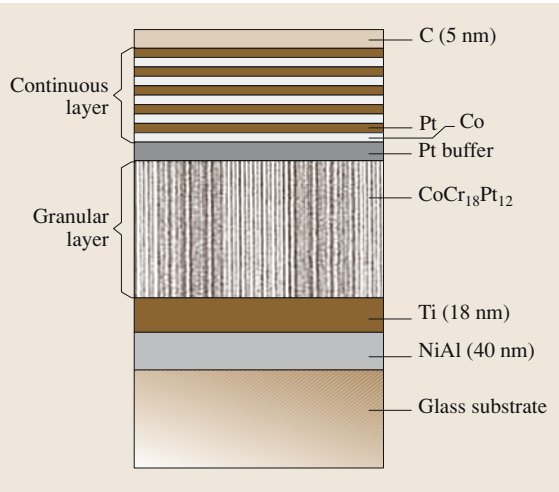


Fig. 51.36 Schematic representation for coupled granular/continuous (CGC) media. (After [51.64])

ular media (good SNR) [51.64, 65, 75]. A cross section of the CGC media is shown in Fig. 51.36 [51.64]. The continuous layer causes an exchange coupling which stabilizes the granular medium; this exchange field is large (≈ 200 kOe) and can result in considerable improvement in thermal stability [51.75]. The initial results for CGC media were encouraging from the standpoint of thermal stability but the high-density SNR was 2 dB less than for the granular media alone. Further optimization of the CGC media using [Co–Pt]_{*n*} multilayers coupled through a Pt buffer layer to a CoCrPt granular layer has shown an improvement in SNR over the granular media alone by 3–5 dB [51.75]. The improvement in SNR is believed to be due to the fact that magnetic recording transitions in the continuous layer are in the form of narrow domain walls and the magnetic interaction between the two layers may result in breaking up the magnetic clusters that form the transitions in the granular medium [51.75].

Magnetic materials used for the soft underlayer (SUL) for a perpendicular recording media must be magnetically soft for low reluctance but also must not exhibit magnetic domains, which can contribute noise to the read signal. The SUL must have a magnetic moment that is larger than that for the pole tip of the write element so that the SUL saturates *after* the pole tip saturates. This condition assures that there is no excess reluctance directly under the pole tip. The SUL must also be thick enough that the reluctance of the path through which the flux travels allows the maximum magnetic field under the gap. Decreasing the reluctance of the flux path also increases the gradient of the magnetic field from the pole, which reduces the length of the transition parameter (51.20) [51.48]. A guide for selecting the SUL thickness (*T*_{SUL}) is

$$T_{\text{SUL}} \approx \frac{A_{\text{pole tip}}}{C_{\text{pole tip}}} \frac{M_{\text{pole tip}}}{M_{\text{SUL}}} \quad (51.28)$$

where *A*_{pole tip} and *C*_{pole tip} are the area and the circumference of the pole tip, respectively, and *M*_{pole tip} and *M*_{SUL} are the saturation magnetizations of the pole tip of the write head and soft underlayer, respectively [51.48]. The effects of changing the thickness of the SUL on the recording parameters of a probe head are shown in Table 51.10 [51.48], where it is seen that the thickness and hence reluctance are key to head performance. Candidate materials for the SUL are listed

Table 51.10 Effect of the thickness of the FeAlN soft underlayer (SUL) on recording head parameters (after [51.48])

Thickness of FeAlN SUL ($4\pi M_s = 20$ kG)	0.1 μm	0.3 μm
Maximum magnetic field (kOe) at recording layer at saturation current (mA) (in parentheses)	16.3 (90)	17.7 (70)
Maximum head field gradient (Oe/nm)	175	550

Table 51.11 Magnetic materials for soft underlayers (SUL)

Material	B_s (kG) ($T \times 10^{-1}$)	Relative Permeability	Comments	Reference
Ni ₈₀ Fe ₂₀ (Permalloy)	10	1500–2000	Insufficient magnetic moment for SUL. Field gradient decreases at maximum field	[51.48]
Sendust (FeSiAl)	10	8000	Insufficient magnetic moment for SUL	[51.4]
FeAlN	20	2000	Sensitive to process conditions; believed to be less useful than FeCoB	[51.48]
FeCoB	20–24	200–240	Annealed to reduce stress	[51.76,77]
CoZrNb	14		Lower magnetostriction than FeCoB	[51.77]
NiFe ₁₇ Mo ₄				[51.78]
[Ni–Fe/Si] _n			Multilayers	[51.79]
IrMn/CoZrNb	14		IrMn antiferromagnetic film to reduce spike noise due to domain walls	[51.80]

Table 51.12 Parameters for perpendicular recording demonstrations

Demonstration	52.5 Gb/in ²	≈ 100 Gb/in ²
media		
Type	CoCrPt	CoCrPt
M_s (emu/cm ³)	250	330
H_c (kOe)	2.6	4.94
K_u (erg/cm ³)	$H_k = 10$ kOe	1.82×10^6
S	0.98	0.92
$\langle D \rangle$ (nm)	–	11.9
Underlayer	FeTaC (16 kG) 400 nm	–
Head		
Turns	–	1
Fly height (nm)		6.4
Magnetic spacing (nm)	20	–
Pole material	FeNi (16 kG)	–
Pole width (nm)	250	105
Pole thickness (nm)	400	105
Write width (nm)	≈ 250	180
Shield–shield spacing (nm)	80	70
Read width (nm)	200	100
Component performance		
Sensitivity (mV/μm)	3	25
PW_{50}		70
Channel performance		
Channel type	Simulator (code rate = 32/33)	Software (code rate = 16/17)
Bit density (kb/in)	590	650
Track density (ktr/in)	88.9	143
Areal density (user/channel) (Gb/in ²)	52.5 (user)	84 (user) 93 (channel)
On-track bit error rate	10^{-5}	6.7×10^{-5}

in Table 51.11. A major issue with the soft underlayer is the motion of domain walls driven by the magnetic field from the write head which can lead to *spike* noise.

The last entry in Table 51.11 is for an antiferromagnetic pinning layer to prevent the motion of such domain walls.

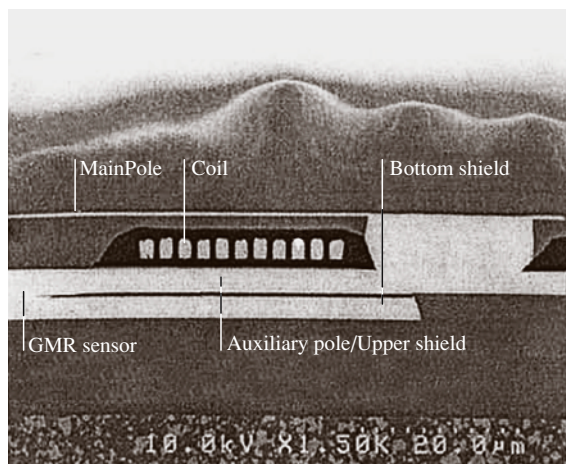


Fig. 51.37 Cross-sectional view of a perpendicular head with a GMR sensor including the bottom and upper shields. The upper shield also serves as the auxiliary pole. (After [51.81])

Perpendicular Recording Demonstrations

Perpendicular recording has now been used in commercial disk drives. The first was the Toshiba 40 GB, 1.8 inch disk drive with an areal density of approximately 75 Gb/in², while more recently the capacity has been increased to 80 GB and the areal density to 133 Gb/in² (Toshiba 2006). Seagate has shipped 2.5 inch notebook drives with a capacity of up to 160 GB and an areal density of 132 Gb/in² (Seagate 2006). An earlier technology demonstration with an areal density of 52.5 Gb/in² was reported in 2001 [51.81] and one for approximately 100 Gb/in² in 2003 [51.82]. Recently, Hitachi Global Storage Technologies has reported an areal density of 230 Gb/in² using perpendicular magnetic recording and a tunneling current-perpendicular-to-the-plane (CPP) GMR read head (to be discussed) [51.83]. We show in Table 51.12 the parameters used for the two areal density demonstrations. In both cases a pole write head and a spin valve read head were used. The parameters for the ≈ 100 Gb/in² demonstration are close to those required for the ultra-high densities (> 100 Gb/in²) as described in Table 51.9, except that the coercivity is about one half that specified. A cross section of the write and read elements for the perpendicular recording head used in the 52.5-Gb/in² demonstration is shown in Fig. 51.37 [51.81]. The pole of the write head is identified as the “main pole” and the thick return pole and upper shield of the GMR sensor are identified as the “auxiliary pole/upper shield”. The read signal versus density (roll-off curve) for the 52.5-Gb/in² demon-

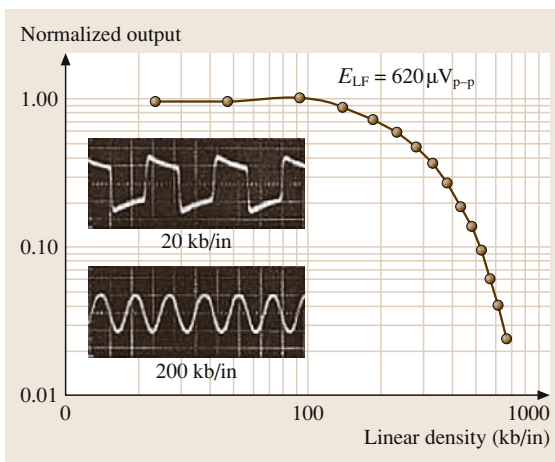


Fig. 51.38 Normalized voltage amplitude for a sequence of read signals from a perpendicular recording system versus linear density in kb/in. (After [51.81])

stration is shown in Fig. 51.38 and results from the increasing interference of positive and negative step pulses in a long string of transitions. At low densities a step response is observed in perpendicular recording since the GMR sensor detects the vertical component of the stray magnetization that reverses at the transition and at higher densities the response appears as a sinusoid. The decay in the signal at the low-density response is due to the insufficient bandwidth of the preamplifier. The read signal for isolated transitions can be described by the response function

$$h(t) = A \tanh \left(\frac{\ln 3}{T_{50}} t \right), \quad (51.29)$$

where A is one half of the signal amplitude and T_{50} is the time for the signal to rise from $-A/2$ to $A/2$. The step signals can be detected either by differentiating the signal to produce pulses, as with longitudinal recording, or with equalizer circuits that directly handle the step response. The channel performance as realized with simulator channels is listed in Table 51.12 and approximates the requirements of a practical system.

Seagate technology reported an areal density of 100 Gb/in² in November 2002 [51.84]. The linear density was 700 kb/in and the track density was 143 ktracks/in. Seagate has also reported an areal density of 170 Gb/in². The linear density was 1050 kb/in and the track density was 161 ktracks/in with a data rate of 483 Mb/s. However, in neither case was this areal density reported in a disk-drive product but rather was measured using a single disk and single head [51.85].

51.2 Magnetic Random-Access Memory

The application of the giant-magnetoresistive effect to recording heads that we have discussed assumed that the sense current is in the plane of the recording film. This is not the only possibility and devices with the current perpendicular to the plane (CPP) have been proposed both for magnetic recording read heads but also for fast magnetic random-access memory (MRAM). The CPP technology that shows the most promise for both the magnetic recording and memory application is the magnetic tunnel junction (MTJ). The MTJ involves a pinned and a free ferromagnetic film spaced by a thin *insulating* film, usually Al_2O_3 . Figure 51.39 shows the schematic of an MTJ used for the MRAM application, including orthogonal bit and word or digit lines and the stack of films comprising a basic MTJ [51.86, 87]. Figure 51.40 shows a more detailed stack of films for an MTJ, including synthetic antiferromagnetic films, antiferromagnetic pinning films and the free film. A typical MTJ has the following thickness values for the key films: antiferromagnet: 300 Å, $\text{Co}_{90}\text{Fe}_{10}$ ferromagnet: 20 Å, ruthenium: 8 Å, Al_2O_3 : 10–15 Å, $\text{Co}_{90}\text{Fe}_{10}$ ferromagnet: 20 Å. As for the spin valve, the use of $\text{Co}_{90}\text{Fe}_{10}$ films for the free layer enhances $\Delta R/R$ [51.20]. Figure 51.40 also shows $\Delta R/R$ for a typical MTJ (for a bit area of $0.6 \times 1.2 \mu\text{m}$) versus magnetic field with the multiple curves being taken over a 6-inch wafer: the maximum change is greater than 48% [51.87]. The change in resistance with orientation of the free layer relative to the pinned layer is significantly larger than for GMR technology, but a thin insulating layer between the ferromagnetic films is required. The memory architecture that has evolved has a single MOS transistor coupled to the MTJ and the word line. The current in the word line supplies the magnetic field along the easy axis of the free film and is insulated from the MTJ (Fig. 51.39). Switching of a soft-magnetic film with a magnetic field applied along the direction of an induced uniaxial anisotropy (the easy axis) and simultaneously along a direction orthogonal to the easy axis (the hard axis) has been discussed (e.g., [51.4]). The switching is described by a switching asteroid given by the following equation and shown in Fig. 51.41 (in the figure the easy and hard magnetic fields are normalized with respect to the anisotropy field H_k).

$$(H_{\text{easy}})^{2/3} + (H_{\text{hard}})^{2/3} = (H_k)^{2/3}$$

$$H_k = 2 \frac{K_u}{M}$$

Switching of the state of the magnetization will occur whenever the combination of easy- and hard-axis fields

is on the outside of the asteroid. The switching field can be significantly reduced if magnetic fields are applied along both the easy- and hard-axis directions and an optimum is to have the fields equal. The bit line is used to supply the hard-axis magnetic field and the word line the easy-axis field. For writing a bit the sense transistor is turned off and current pulses are applied along both orthogonal conductors, generating a magnetic field at the free film sufficient in amplitude to switch the magnetization of the free film. For reading a bit, the sense transistor is turned on by the word line and pulses of current are driven along the bit line (with the amplitude of the magnetic field insufficient to switch the free film) and the current is either high when the ΔR is low, corresponding to say a 1 or the current amplitude is low when ΔR is high, corresponding to a 0. Without the sense transistor there would be many *sneak paths* for the current through other MTJs, reducing the available current for reading the selected bit. The MTJ memory technology shows considerable promise for replacing semiconductor SRAM and DRAM memories, primarily because of the nonvolatility of the memory in the MRAM technology. The speed for reading data is comparable to that of SRAM and the density is comparable to that for DRAM.

The physics of the MTJ is different from that for GMR [51.88]. Ferromagnetic films have two types of

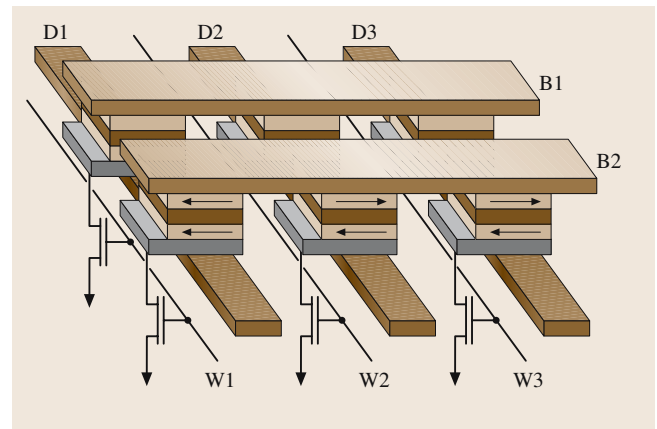


Fig. 51.39 Schematic of MRAM cross-point architecture with bits between orthogonal conductors and each cell defined by one MTJ and one transistor. The top lines in contact with the top electrode of the bits provides hard-axis fields, while the bottom lines are isolated and provide easy-axis fields. Turning on the transistor provides a current path so that the corresponding bit can be sensed. (After [51.86])

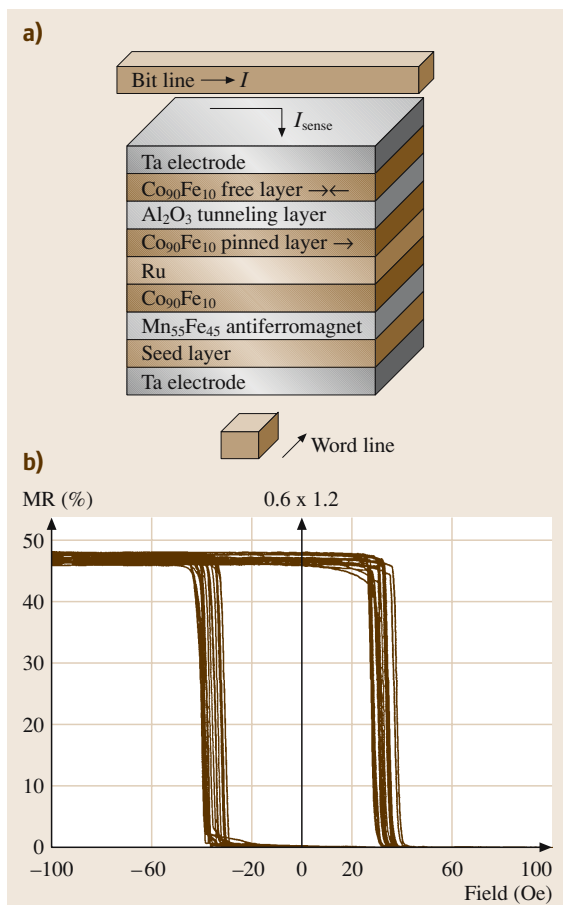


Fig. 51.40 (a) Stack of films comprising a typical MTJ. The arrows show the direction of the magnetizations in the two coupled films. (b) Hysteresis curves of $\Delta R/R$ versus magnetic field for $0.6 \times 1.2 \mu\text{m}$ bits on a 6-in wafer. (After [51.87])

electrons: those polarized parallel to the magnetization (\uparrow) and those polarized antiparallel to the magnetization (\downarrow). A polarization ratio (P) for a ferromagnetic material can be defined as

$$P = \frac{D(E_F)^\uparrow - D(E_F)^\downarrow}{D(E_F)^\uparrow + D(E_F)^\downarrow},$$

where $D(E_F)^\uparrow$ and $D(E_F)^\downarrow$ are the densities of states for electrons at the Fermi energy for the spin-up and spin-down electrons, respectively. The magnetoresistance ratio of two films coupled by an insulator is [51.89]

$$\frac{\Delta R}{R} = \frac{R_a - R_p}{R_p} = \frac{2P_1 P_2}{1 - P_1 P_2}, \quad (51.30)$$

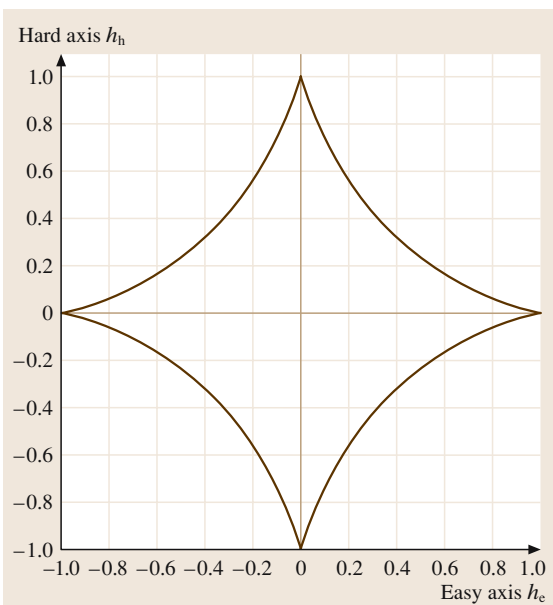


Fig. 51.41 Switching asteroide plotted with magnetic fields normalized to the induced anisotropy field ($2K_u/M$). The ferromagnetic film will switch for any combination of easy- and hard-axis fields on the outside of the asteroide

where R_a and R_p are the resistances of the coupled films in the antiparallel and parallel configurations, respectively, and the spin polarizations are P_1 and P_2 for the two ferromagnetic films. Values of the spin polarization ratio are: $P_{\text{Co}} \approx 0.35$, $P_{\text{Fe}} \approx 0.44$ and $P_{\text{Permalloy}} \approx 0.3$ [51.88]. With cobalt used as one electrode and Permalloy for the other electrode, the predicted value for $\Delta R/R$ is 24%. Conduction through the insulating film is by quantum-mechanical tunneling, described by a transmission coefficient (T)

$$T = \frac{|\psi_{\text{transmitted}}(x)|^2}{|\psi_{\text{incident}}(x)|^2},$$

where $\psi(x)$ is the wavefunction for the electron plane wave and $|\psi(x)|^2$ is the probability of finding an electron at the position x . The transmission coefficient is a measure of the resistance of the barrier. The transmission coefficient for electrons with energy E tunneling through an insulating film of thickness a with amplitude of the barrier height of V_0 is given by

$$T = T_0 e^{-2\alpha a},$$

where

$$T_0 = \frac{16E(V_0 - E)}{V_0^2},$$

Table 51.13 Parameters for Permalloy (Ni₈₀Fe₂₀)

M_s (T)	K_1 (J/m ³)	A (exchange constant) (10 ⁻¹² J/m)	Critical length l_s (nm)
1.0	500	13	5.7

and

$$\alpha^2 = \frac{2m(V_0 - E)}{\hbar^2}.$$

This result applies when $\alpha a > 1$. A barrier height of 10 eV applies for CoO. The exponential dependence of the transmission coefficient on the thickness of the barrier shows the requirement for a thin barrier. A typical value of the thickness of an Al₂O₃ layer is 10 Å and the resistance changes by a factor of 100 million in changing the thickness from 0 to 10 Å. The reason that the MTJ is more attractive than a spin valve for the memory application is that the MTJ can be made with a resistance of many ohms, while the spin valve is nearly a short circuit. The use of the synthetic antiferromagnet for the MTJ also helps to avoid the superparamagnetic limit for this device, since the films comprising the synthetic antiferromagnet are thicker than the single antiferromagnetic film that was used to pin one of the ferromagnetic electrodes.

A significant problem with the MTJ is the different magnetic states that can be found depending on the size and shape of the bit. A review of an analysis of the states for a rectangular-shaped bit has been published [51.90] and we will quote some of their results to illustrate the problem. The analysis uses finite element analysis to solve the equation of motion for the magnetization (\mathbf{M}) with Gilbert damping in the presence of an effective magnetic field (\mathbf{H}_{eff})

$$\frac{d\mathbf{M}}{dt} = \gamma (\mathbf{M} \times \mathbf{H}_{\text{eff}}) - \frac{\alpha}{M} (\mathbf{M} \times \dot{\mathbf{M}}). \quad (51.31)$$

The effective magnetic field can be derived from the magnetic free-enthalpy density

$$\mathbf{H}_{\text{eff}} = -\delta g / \delta \mathbf{M}$$

and the magnetic free enthalpy per unit volume is given by the sum of terms

$$g = g_{\text{exchange}} + g_{\text{anisotropy}} + g_{\text{demagnetizing-field}} + g_{\text{external-field}}.$$

An example of this analysis is a rectangular film of thickness 20 nm, length 1 μm, width 1 μm with magnetic properties shown in Table 51.13. The exchange constant is $A = s^2 a^2 J_{\text{ex}} N_v / 2$, where s is the spin of the atom, a the lattice constant and N_v is the number of nearest-neighbor atoms per unit volume. The critical length determines the maximum size of the mesh used in the finite element analysis. Figure 51.42 shows three of the six states of the magnetization in the sample. The first two are high-remanence states referred to as S-type and C-type and can be formed if the starting conditions are a single-domain state. The last state shown is a low-remanence state with an off-center vortex. The last state is formed when the initial conditions are two domains along the long axis of the film. The total energies for the three states are shown in Table 51.14 and it is seen that the lowest energy is the low-remanence state. An issue with the MRAM is the possibility of exciting the different states in the storage cell that can then result in a different switching asteroid. Thin films with low-aspect-ratio bit shapes have been predicted to result in reduced propensity to excite the vortex structures [51.91]. Changing the shape of the bit may also assist in minimizing the vortex structures. Software for analyzing these problems is the LLG Micromagnetics SimulatorTM developed by M. R. Scheinfein (llg@dancris.com).

A recent development in MRAM architecture is to pass the write current through the MTJ to heat the element and to lower the threshold for writing [51.92].

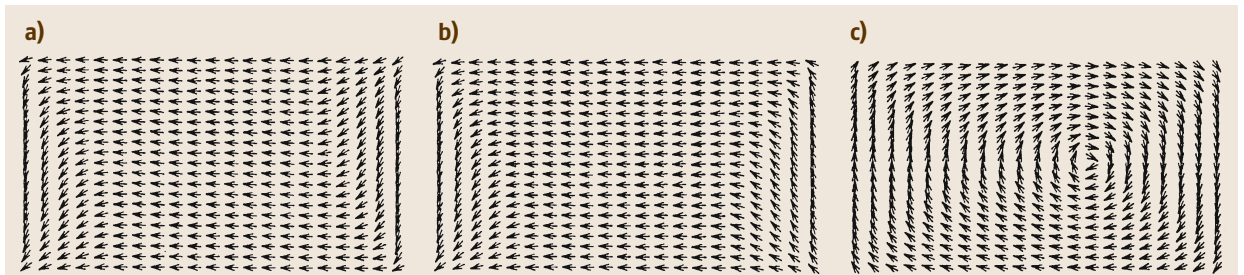


Fig. 51.42a–c Configurations for spins in a Permalloy sample described in Tables 51.13 and 51.14. Figure (a) is an S-state, (b) is a C-state and (c) is an off-center vortex state. (After [51.90])

Table 51.14 Total energies for different spin configurations in the Permalloy film (described in the text and Table 51.13) (after [51.90])

Structure	G (10 ⁻¹⁸ J)
S-state	169 897
C-state	169 890
Shifted vortex state	79 616

This is accomplished by turning on the selection transistor (Fig. 51.39) during the write operation. The advantages of this modification to the MRAM architecture are:

1. since the selection at write is controlled by the temperature, selection errors are reduced,
2. the write threshold is reduced,
3. thermal instability limits are increased by using materials with larger switching fields at the operating temperature [51.92].

51.2.1 Tunneling Magnetoresistive Heads

Tunneling magnetoresistance (TMR) can be used in a magnetic recording read-head application. These TMR heads use the CPP orientation of the sense current and, as with the MRAM, the tunneling is done through a thin insulating barrier. A schematic at the air-bearing surface in the simplest form of such a TMR head is shown as Fig. 51.43 [51.93]. The insulating barrier (again usually fabricated from Al₂O₃ films) is shown on edge and the current flow is in the plane of the paper. Longitudinal bias to put the ferromagnetic TMR electrode films in a single-domain state and the top and bottom shields are shown. The GMR ratio with the TMR effect is large but the resistance × area (RA) product, which is the resistivity times the film thickness, may also be large, resulting in excessive shot noise. Shot noise is intrinsic to tunneling with the direct-current (DC) bias current required for GMR head operation and is the dominant noise source in the TMR head. Figure 51.44 is a plot of the TMR ratio in % versus the RA product for three different TMR structures all made from the basic structure: buffer/PtMn/CoFe/Ru/CoFe/AlO_x/CoFe/NiFe/cap [51.93]. The PtMn is an antiferromagnet, AlO_x is the insulating barrier (prepared by in situ oxidation) and CoFe/NiFe is the second electrode. The data points for TMR versus RA for this structure are labeled “Conv-btm” (conventional bottom) and the data points for the “Smooth-top” structure is for the PtMn layer on top

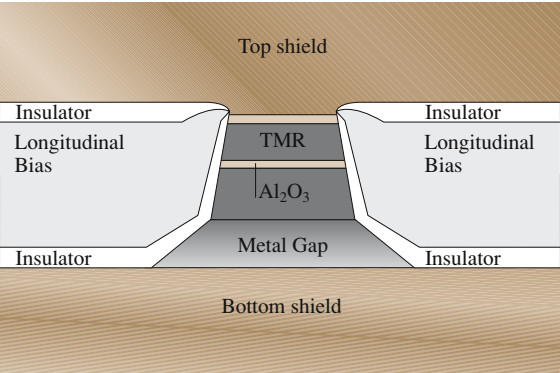


Fig. 51.43 Schematic drawing of the air-bearing surface view of a tunneling magnetoresistive (TMR) head. (After [51.93])

instead of the bottom, while the “Smooth-btm” data points are for a head made with smoother interfaces. The TMR % remains constant at ≈ 20% until a knee, below which the TMR % drops due to pinholes and incomplete insulation coverage. The stack of films is ion-milled and the stripe height is defined by lapping. Magnetic recording results showed the expected large signal ($V_{\text{peak-to-peak}} = 42 \text{ mV}/\mu\text{m}$) with a biasing current of 1 mA but the SNR was only 23.4 dB and was not deemed sufficient to demonstrate the required low error rate [51.93]. A significant problem with the TMR head is that there is a large resistance in parallel with the

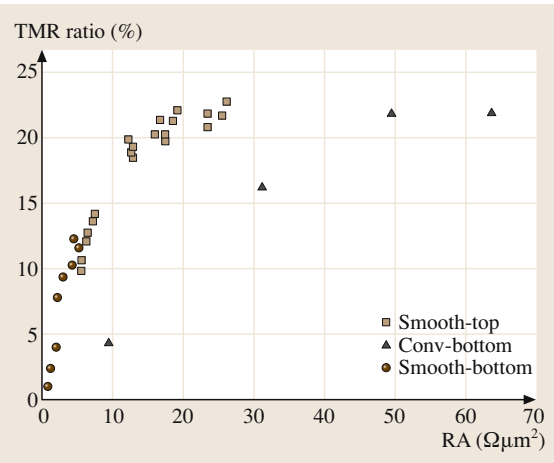


Fig. 51.44 Correlation between TMR ratio and resistance area product (RA) in three structures: top- and bottom-type MTJ with smoother interfaces and bottom-type MTJ with conventional stacking that has relatively rough interfaces. (After [51.93])

large capacitance resulting from the tunneling barrier. This combination results in a low-pass filter and care is required to limit the capacitance. In the example discussed here the cut off linear density is near 580 kfc/in.

Because of the significantly larger value of ΔR for the CPP as opposed to the CIP technology, it is possible that CPP may become the dominant technology for read heads in the future.

51.3 Extraordinary Magnetoresistance (EMR)

An interesting development in read head technology is the discovery that narrow band gap and hence high-mobility semiconductors can be made to function as detectors of magnetic field [51.94]. The concept is based on the Lorentz force exerted on charge carriers by a magnetic field

$$F = q\vec{v} \times \vec{B},$$

here q is the charge, \vec{v} is the velocity of the charge carrier and \vec{B} is the magnetic flux density acting on the charge. Electrons have the highest mobility as compared to holes and are the carriers used. The electrons are driven by an external source similar to the sense current in GMR sensors in a high mobility semiconductor and are deflected by the external magnetic field

into a trajectory and the voltage drop in the semiconductor at low locations is used to sense the magnetic field. The III-IV semiconductor InSb was chosen for the initial work on EMS because of the high mobility of electrons of $60\,000\text{ cm}^2/\text{V s}$. However, a major limitation was the ability to maintain the high mobility in thin films ($< 100\text{ nm}$) required for read head applications and this was addressed by forming a thin-film quantum well heterostructure [51.94]. An advantage of such sensors is the lack of any magnetic noise since there are no magnetic materials involved in the head structure. No experimental results of such structures in read head applications has been reported. A review of EMR as magnetic field nanosensors has been reported by S. Solin [51.95].

51.4 Summary

It was the purpose of this chapter to review the current status of the technology of magnetic recording as used in disk drives. The emphasis was on the magnetic materials used in the application and on some of the technical problems that may limit the increase in areal density of the devices. The new technology of MRAM, which has evolved from the magnetic recording application, was also reviewed. It was found that a wide range of magnetic materials is essential for the advance of magnetic recording and the MRAM technology, including high-magnetization soft-magnetic materials for write heads, new antiferromagnetic alloys with high

blocking temperatures and low susceptibility to corrosion for pinning films in giant-magnetoresistive sensors and new ferromagnetic alloys with large values of giant magnetoresistance. A significant limitation to magnetic recording was found to be the superparamagnetic effect and new advances in multilayer ferromagnetic films to reduce the impact of the effect, but also to allow high-density recording, were discussed. Perpendicular recording as an alternative to longitudinal recording was reviewed and it was concluded that this technology will be the dominant recording technology in the future.

References

- | | |
|---|---|
| <p>51.1 E. Grochowski: IBM Almaden Research Laboratories (http://www.storage.ibm.com/technology/grochows/grocho01.htm) (2003)</p> <p>51.2 R. L. Comstock: J. Mater. Sci. Mater. Electron. 13, 509 (2002)</p> <p>51.3 C. D. Mee, E. D. Daniel, M. H. Clark: <i>Magnetic Recording: The First 100 Years</i> (Wiley, New York 1998)</p> | <p>51.4 R. L. Comstock: <i>Introduction to Magnetism and Magnetic Recording</i> (Wiley, New York 1999)</p> <p>51.5 R. L. Comstock: <i>Data storage in rigid disks</i>, 2nd edn., ed. by C. Mee, E. Daniels (McGraw-Hill, New York 1996)</p> <p>51.6 R. Bozorth: <i>Ferromagnetism</i> (Van Nostrand, New York 1951)(reprinted by IEEE, New York 1993)</p> |
|---|---|

- 51.7 X. Liu, G. Zangart, M. Shamsuzzoha: J. Electrochem. Soc. **150** (3), C159 (2003)
- 51.8 N. Robertson, B. Hu, C. Tsang: IEEE Trans. Magn. **33**, 2818 (1997)
- 51.9 X. Liu, G. Zangari: IEEE Trans. Magn. **37**, 1764 (2001)
- 51.10 T. Osaka, T. Yokoshima, T. Nakanishi: IEEE Trans. Magn. **37**, 1761 (2001)
- 51.11 Y. Liu, V. Harris, M. Kryder: IEEE Trans. Magn. **37**, 1779–1782 (2001)
- 51.12 Y. Ding, S. Byeon, C. Alexander Jr.: IEEE Trans. Magn. **37**, 1776 (2001)
- 51.13 R. E. Jones Jr.: IBM Disk Storage Technol. **GA-26-1665-0**, 6 (1980)
- 51.14 P. Andricacos, L. Romankiw: Magnetically soft materials in data storage: Their properties and electrochemistry. In: *Adv. Electrochem. Sci. Eng.*, Vol. 3, ed. by H. Gerischer, C. Tobias (VCH, Weinheim 1990)
- 51.15 R. L. Comstock, M. Williams: AIP Conf. Proc. Magn. Mater. **5**, 738 (1971)
- 51.16 T. Dinan: Electrolytics, Electrolytics Inc., San Jose Ca, private communication
- 51.17 G. Bertero, S. Malhotra, B. Bian, J. Tsoi, M. Avenell, D. Wachenschwanz, T. Yamashita: IEEE Trans. Magn. **39**, 651 (2003)
- 51.18 S. Byeon, F. Liu, G. Mankey: IEEE Trans. Magn. **37**, 1770 (2001)
- 51.19 M. Baibach, J. Broto, A. Fert, F. Nguyen Van Dan, F. Petroff, P. Etienne, G. Creuzet, A. Fredrick, J. Chazelas: Phys. Rev. Lett. **61**, 2472 (1988)
- 51.20 H. Kanai, K. Yamada, K. Aoshima, Y. Ohtsuka, J. Kane, M. Kanamine, J. Toda, Y. Mizoshita: IEEE Trans. Magn. **32**, 3368 (1996)
- 51.21 R. Wood, M. Williams, J. Hong: IEEE Trans. Magn. **26**, 2954 (1990)
- 51.22 B. Dieny, V. Speriosu, S. Parkin, P. Baumgart, D. Wilhoit: J. Appl. Phys. **69**, 4774 (1991)
- 51.23 S. Parkin: Appl. Phys. Lett. **61**, 1358 (1992)
- 51.24 H. Bertram, H. Zhou, R. Gustafson: IEEE Trans. Magn. **34**, 1846 (1998)
- 51.25 T. Lin, C. Tsang, R. Fontana, J. Howard: IEEE Trans. Magn. **32**, 2585 (1995)
- 51.26 A. Devasahayam, M. Kryder: IEEE Trans. Magn. **35**, 649 (1999)
- 51.27 Y. Hamakawa, M. Komuro, K. Watanabi, H. Hoshiya, T. Okada, K. Nakamoto, Y. Suzuki, M. Fuyama, H. Fukui: IEEE Trans. Magn. **35**, 677 (1999)
- 51.28 H. Kishi, Y. Kitade, Y. Miyaki, A. Tanaka, K. Kobayashi: IEEE Trans. Magn. **32**, 3380 (1996)
- 51.29 S. Parkin, D. Mauri: Phys. Rev. B **44**, 7131 (1991)
- 51.30 J. Mathon: Contemp. Phys. **32**, 143 (1999)
- 51.31 S. O. Kasap: *Principles of Electronic Materials and Devices*, 2nd edn. (McGraw-Hill, New York 2002)
- 51.32 I. A. Campbell, A. Fert: Transport properties of ferromagnets. In: *Ferromagnetic Materials*, Vol. 3, ed. by E. P. Wohlfarth (North-Holland, Amsterdam 1984) p. 3
- 51.33 R. C. O'Handley: *Modern Magnetic Materials Principles and Applications* (Wiley, New York 2000)
- 51.34 R. L. White: IEEE Trans. Magn. **28**, 2482 (1992)
- 51.35 M. Doerner, X. Bian, M. Madison, K. Tang, Q. Peng, A. Polcyn, T. Arnoldussen, M. Toney, M. Mirzamaani, K. Takano, E. Fullerton, D. Margulies, M. Schabes, K. Rubin, M. Pinarbasi, S. Yuan, M. Parker, D. Weller: IEEE Trans. Magn. **37**, 1052 (2001)
- 51.36 T. Arnoldussen: IEEE Trans. Magn. **34**, 1851 (1998)
- 51.37 G. Choe, J. Zhou, B. Demczyk, M. Yu, M. Zheng, R. Weng, A. Chekanov, K. Johnson, F. Liu, K. Stoev: IEEE Trans. Magn. **39**, 633 (2003)
- 51.38 M. Madison, T. Arnoldussen, M. Pinarbasi, T. Chang, M. Parker, J. Li, S. Duan, X. Bian, M. Mirzamaani, R. Payne, C. Fox, R. Wang: IEEE Trans. Magn. **35**, 695 (1999)
- 51.39 D. Weller, A. Moser: IEEE Trans. Magn. **35**, 4423 (1999)
- 51.40 E. Abarra, B. Acharya, A. Inomata, I. Okamoto: IEEE Trans. Magn. **37**, 1426 (2001)
- 51.41 E. Abarra, A. Inomata, H. Sato, I. Okamoto, Y. Mizoshita: Appl. Phys. Lett. **77**, 2581 (2000)
- 51.42 E. Fullerton, D. Margulies, N. Schabes, M. Carey, B. Gurney, A. Moser, M. Best, G. Zeltzer, K. Rubin, H. Rosen: Appl. Phys. Lett. **77**, 3806 (2000)
- 51.43 A. Inomata, B. Acharya, E. Abarra, A. Ajan, D. Hasegawa, I. Okamoto: J. Appl. Phys. **91**, 7671 (2002)
- 51.44 M. Schabes, E. Fullerton, D. Margulies: IEEE Trans. Magn. **37**, 1432 (2001)
- 51.45 IBM Research News (http://www.research.ibm.com/resources/news/20010518_whitepaper.shtml) (2001)
- 51.46 Z. Zhang, Y. Feng, T. Clinton, G. Badran, N. Yeh, G. Tarnopolsky, E. Girt, M. Munteanu, S. Harkness, H. Richter, T. Nolan, R. Ranjan, S. Hwang, G. Rauch, M. Ghaly, D. Larson, E. Singleton, V. Vasko, J. Ho, F. Stageberg, V. Kong, K. Duxstad, S. Slade: IEEE Trans. Magn. **38**, 1861 (2002)
- 51.47 G. Choe, J. Zhou, R. Weng, K. Johnson: J. Appl. Phys. **91**, 7665 (2002)
- 51.48 S. Khizroev, Y. Lui, K. Mountfield, M. Kryder, D. Litvinov: J. Magn. Magn. Mater. **246**, 335 (2002)
- 51.49 M. Mallory, A. Torabi, M. Benaki: IEEE Trans. Magn. **38**, 1719 (2002)
- 51.50 M. L. Williams: Conf. Proc. IDEMA Symposium Perpendicular Recording, (2004)
- 51.51 W. Xia, H. Aoi, H. Muraoka, Y. Nakamura: IEEE Trans. Magn. **40**, 2365 (2004)
- 51.52 T. Okada, H. Kimura, I. Nunokawa, N. Yoshida, K. Etoh, M. Fuyama: IEEE Trans. Magn. **40**, 290 (2004)
- 51.53 N. Honda, K. Ouchi, S. Iwasaki: IEEE Trans. Magn. **38**, 1615 (2000)
- 51.54 K. Ouchi, N. Honda: IEEE Trans. Magn. **36**, 16 (2002)
- 51.55 N. Honda, S. Yanase, K. Ouchi, S. Iwasaki: J. Appl. Phys. **85**, 6130 (1999)

- 51.56 I. Lee, H. Ryu, H. Lee, T. Lee: J. Appl. Phys. **85**, 6133 (1999)
- 51.57 T. Kertoku, J. Ariake, N. Honda, K. Ouchi: J. Magn. Magn. Mater. **235**, 34 (2001)
- 51.58 L. Wu, S. Yanase, N. Honda, K. Ouchi: J. Magn. Soc. Jpn. **21**, 301 (1997)
- 51.59 X. Qi, B. Stadler, R. Victora, J. Judy, O. Hellwig, N. Supper: IEEE Trans. Magn. **40**, 2476 (2004)
- 51.60 T. Suzuki, N. Honda, K. Ouchi: J. Magn. Soc. Jpn. **21-S2**, 177 (1997)
- 51.61 G. Bertero, D. Wachenschwanz, S. Malhotra, S. Velu, B. Bian, D. Stafford, Y. Wu, T. Yamashita, S. Wang: IEEE Trans. Magn. **38**, 1627 (2002)
- 51.62 T. Oikawa, M. Nakamura, H. Uwazumi, T. Shimatsu, H. Muraoka, Y. Nakamura: IEEE Trans. Magn. **38**, 1976 (2002)
- 51.63 H. Uwazumi, K. Enomoto, Y. Sakai, S. Takenoiri, T. Oikawa, S. Watanabe: IEEE Trans. Magn. **39**, 1914 (2003)
- 51.64 Y. Sonobe, D. Weller, Y. Ikeda, K. Takano, M. Schabes, G. Zeltzer, H. Do, R. Yeu, M. Best: J. Magn. Magn. Mater. **235**, 424 (2001)
- 51.65 Y. Sonobe, K. Miura, Y. Nakamura, K. Takano, H. Do, A. Mosher, B. Yen, Y. Ikeda, N. Supper: J. Appl. Phys. **91**, 8055 (2002)
- 51.66 H. Muraoka, Y. Sonobe, K. Muira, A. Goodman, Y. Nakamura: IEEE Trans. Magn. **38**, 1632 (2002)
- 51.67 Y. Sonobe, H. Muraoka, K. Miura, Y. Nakamura, K. Takano, A. Moser, H. Do, B. Yen, Y. Ikeda, N. Supper, H. Weresin: IEEE Trans. Magn. **38**, 2006 (2002)
- 51.68 M. Williams, C. Rettner, K. Takano, W. Weresin: IEEE Trans. Magn. **38**, 1643 (2002)
- 51.69 R. M. White: *Introduction to Magnetic Recording* (IEEE, New York 1985)
- 51.70 K. Z. Gao, H. N. Bertram: IEEE Trans. Magn. **38**, 3675 (2002)
- 51.71 J. Judy: J. Magn. Magn. Mater. **235**, 235 (2001)
- 51.72 E. Velu, S. Malhotra, G. Bertero, D. Wachenschwanz: IEEE Trans. Magn. **39**, 668 (2003)
- 51.73 S. Oikawa, A. Takeo, T. Hikosaka, Y. Tanaka: IEEE Trans. Magn. **36**, 2393 (2000)
- 51.74 M. Zheng, G. Choe, A. Chekanov, B. Demczyk, B. Acharya, K. Johnson: IEEE Trans. Magn. **39**, 1919 (2003)
- 51.75 K. Miura, H. Muraoka, Y. Sonobe, Y. Nakamura: IEEE Trans. Magn. **38**, 2054 (2002)
- 51.76 C. Brucker, T. Nolan, B. Lu, Y. Kubota, M. Plumer, P. Lu, R. Cronch, C. Chang, D. Chen, J. Chen, R. Michel, G. Parker, N. Tabat: IEEE Trans. Magn. **39**, 673 (2003)
- 51.77 C. Chang, M. Plumer, C. Brucker, J. Chen, R. Ranjan, J. van Elk, J. Yu, D. Karns, Y. Kubota, G. Ju, D. Weller: IEEE Trans. Magn. **38**, 1637 (2002)
- 51.78 M. Zheng, G. Choe, K. Johnson, L. Gao, S. Liou: IEEE Trans. Magn. **38**, 1979 (2002)
- 51.79 S. Kong, T. Okamoto, K. Kim, S. Nakagawa: IEEE Trans. Magn. **38**, 1982 (2002)
- 51.80 S. Takenoiri, K. Enomoto, Y. Sakai, S. Watanabe: IEEE Trans. Magn. **38**, 1991 (2002)
- 51.81 H. Takano, Y. Nishida, A. Kuroda, H. Sawaguchi, Y. Hosoe, T. Kawabe, H. Aoi, H. Muraoka, Y. Nakamura, K. Ouchi: J. Magn. Magn. Mater. **235**, 241 (2001)
- 51.82 W. Eppler, A. Sunder, D. Karns, E. Kurtas, G. Ju, X. Wu, P. van der Heijden, Y. Kubota, H. Zhou, C.-H. Chang: IEEE Trans. Magn. **39**, 663 (2003)
- 51.83 Hitachi Global Storage Technologies, announcement on the website; www.hitachigst.com/hdd/research/recording_head/pr (March 2005)
- 51.84 Seagate press release: seagate.com (2002)
- 51.85 E. Murdock, J. van Ek: Lake arrowhead workshop (2003)
- 51.86 S. Parkin, K. Roche, M. Samant, P. Rice, R. Beyers, R. Scheuerlein, E. O'Sullivan, S. Brown, J. Bucchigano, D. Abraham, Y. Lu, M. Rooks, P. Trouilloud, R. Wanner, W. Gallagher: J. Appl. Phys. **85**, 5828 (1999)
- 51.87 S. Tehrani, B. Engel, J. Slaughter, E. Chen, M. DeHerra, M. Durlam, P. Naji, R. Whig, J. Janesky, J. Calder: IEEE Trans. Magn. **36**, 2752 (2000). The curve of magnetoresistance % is from a private communication with Mark deHerrera from Motorola
- 51.88 R. Meservey, P. Tedrow: Phys. Rep. **238**, 175 (1994)
- 51.89 M. Julliere: Phys. Lett. **54A**, 225 (1975)
- 51.90 H. Kronmüller, R. Hertel: J. Magn. Magn. Mater. **215-216**, 11 (2000)
- 51.91 K. Lee, W. Park, T. Kim: IEEE Trans. Magn. **39**, 2842 (2003)
- 51.92 I. Prejbeanu, W. Kula, K. Ounadjela, R. Sousa, O. Redon, B. Dieny, J. Nozieres: IEEE Trans. Magn. **40**, 2625 (2004)
- 51.93 S. Araki, K. Sato, T. Kagami, S. Saruki, T. Uesugi, N. Kasahara, T. Kuwashima, N. Ohta, J. Sun, K. Nagi, S. Li, N. Hachisuka, H. Hatata, N. Kagotani, N. Takahashi, K. Ueda, M. Matsuzaki: IEEE Trans. Magn. **38**, 72 (2002)
- 51.94 S. Solin, D. Hines, A. Rowe, J. Tsai, Y. Pashkin, S. Chung, N. Goel, M. Santos: Appl. Phys. Lett. **80**, 4012 (2002)
- 51.95 S. Solin: Sci. Am. Mag. **291**, 71 (2004)

52. High-Temperature Superconductors

The discovery by J. G. Bednorz and K. A. Müller in 1986 that the superconducting state can exist in oxides at temperatures above 30 K stimulated research in the field of superconductivity and opened a new field of research. Within a few years a large number of cuprate superconductors with transition temperatures well above the boiling point of liquid nitrogen have been found. In this chapter an overview of the major families of high-temperature superconductors and their physical properties is presented.

Starting from the well-known characteristics of conventional superconductors, described in Sect. 52.1, the new phenomena observed in high-temperature superconductors are considered. The complexity of the physical properties of the cuprate superconductors is closely related to the fact that these materials are close to a metal–insulator transition. In Sects. 52.2 and 52.3 the crystal structures, the general trends for the critical temperatures, the anisotropy of the physical properties and the factors limiting the transport critical current density are discussed. Because of their importance in the field of electronics some features of thin films are presented in Sect. 52.4.

The binary compound MgB_2 is strictly speaking not a high-temperature superconductor. Nevertheless, an overview of the physical properties of this interesting metallic superconductor, charac-

52.1	The Superconducting State	1195
52.1.1	Characteristic Properties of Superconductors	1195
52.1.2	Superconductor Electrodynamics	1197
52.1.3	Superconductivity: A Macroscopic Quantum Phenomenon	1198
52.1.4	Type II Superconductors	1200
52.2	Cuprate High-T_c Superconductors: An Overview	1202
52.2.1	Major Families of Cuprate Superconductors	1202
52.2.2	Generic Phase Diagram of Cuprate Superconductors	1202
52.2.3	Crystal Structures	1204
52.2.4	Critical Temperatures	1205
52.3	Physical Properties of Cuprate Superconductors	1207
52.3.1	Anisotropic Superconductors	1207
52.3.2	Irreversibility Line	1208
52.3.3	Limitations of the Transport Critical Current ..	1209
52.4	Superconducting Films	1212
52.5	The Special Case of MgB_2	1214
52.6	Summary	1216
	References	1216

terised by a transition temperature as high as 39 K, has been included in the present chapter (Sect. 52.5).

After the successful development of helium liquefaction techniques in the laboratory of Heike Kammerlingh Onnes at the University of Leiden, temperatures down to about 1 K became accessible for further research [52.1]. One of the first aspects to be studied was the electrical resistivity of pure metals at very low temperatures. The nearly temperature-independent residual resistivity of platinum and gold was found to decrease with increasing purity of the metals. In 1911 Kammerlingh Onnes found that, in contrast to the behaviour of platinum and gold, the electrical resistivity of mercury drops to an unmeasurably small value at

a temperature of 4.2 K, as shown in Fig. 52.1. The remarkable phenomenon of superconductivity had been discovered [52.1].

In addition to zero resistance, the superconducting state is characterised by perfect diamagnetism. *W. Meissner* and *R. Ochsenfeld* found in 1933 that, as soon as the superconducting state is reached, a magnetic field is expelled from the interior of a superconductor cooled in the presence of a static magnetic field [52.2]. In contrast to the shielding of a magnetic field applied to a material in the superconducting state, field expulsion cannot be explained by perfect conductivity. This Meissner effect

Table 52.1 Milestones in the history of superconductivity

1911	H. Kammerlingh Onnes discovers superconductivity in Hg [52.1]
1933	W. Meissner and R. Ochsenfeld find perfect diamagnetism for the superconducting state [52.2]
1935	London theory of the superconductor electrodynamics [52.3]
1957	J. Bardeen, L. N. Cooper and J. R. Schrieffer develop a quantum theory of superconductivity (BCS theory) [52.4, 5]
1960	I. Giaever measures the energy gap by electron tunnelling [52.6, 7]
1961	Experimental confirmation of flux quantisation indicating the existence of Cooper pairs [52.8, 9]
1962	B. D. Josephson predicts Cooper pair tunnelling between two superconductors separated by a thin insulating oxide layer [52.10]
1965	Observation of superconducting quantum interference between Josephson junctions in multiply connected superconducting circuits [52.11]
1986	J. G. Bednorz and K. A. Müller discover high-temperature superconductivity in the La-Ba-Cu-O system [52.12]
1993	Critical temperature of 135 K in HgBa ₂ Ca ₂ Cu ₃ O _{8+δ} [52.13, 14]
2001	J. Akimitsu et al. discover superconductivity at 39 K in MgB ₂ [52.15]

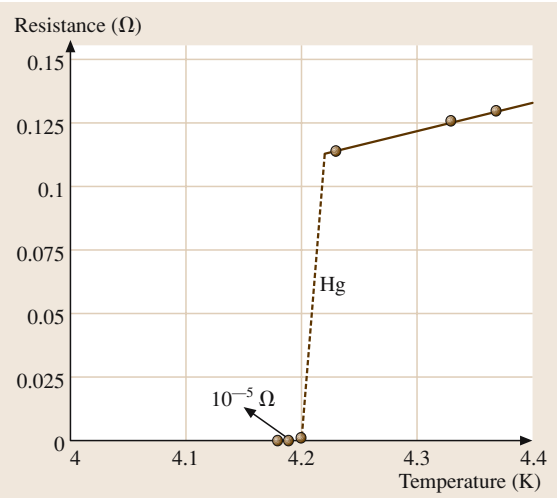


Fig. 52.1 Resistance–temperature plot for mercury obtained by Heike Kammerlingh Onnes

is therefore indeed an additional characteristic property of superconductors.

The discoveries listed in Table 52.1 provide an overview of the history of superconductivity. The development of the highest known critical temperatures in metals and oxides is shown in Fig. 52.2. Between 1911 and 1974, the critical temperatures T_c of metallic superconductors steadily increased from 4.2 K in mercury up to 23.2 K in sputtered Nb₃Ge films [52.16, 17]. Nb₃Ge held the record for the critical temperature in metallic

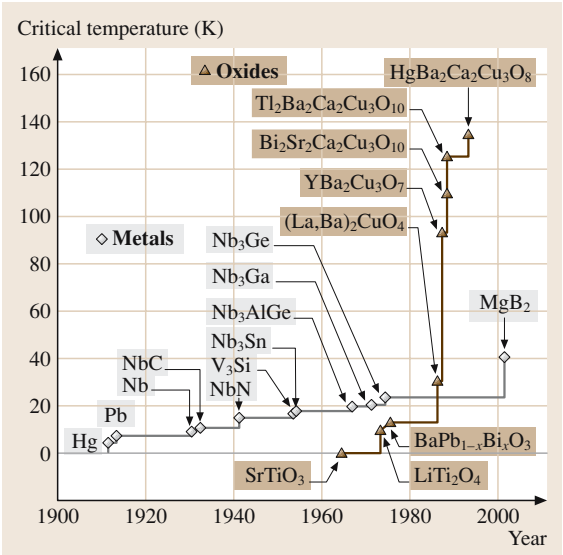


Fig. 52.2 The highest known critical temperatures of metallic and oxide superconductors

superconductors until the unexpected discovery of superconductivity at 39 K in the intermetallic compound MgB₂ [52.15].

The first superconducting oxide SrTiO₃ characterised by a transition temperature as low as 0.25 K was discovered in 1964 [52.18]. A remarkably higher critical temperature of 13 K was found for the perovskite BaPb_{1-x}Bi_xO₃ in 1975 [52.19]. These results

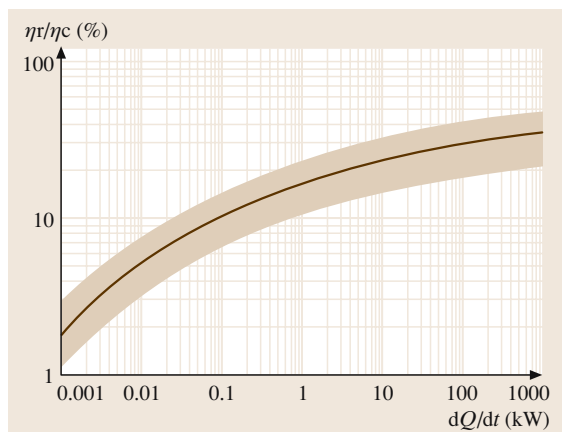


Fig. 52.3 The efficiency of real refrigerators normalised to the Carnot efficiency is typically in the hatched area. The ratio of η_r/η_C increases with increasing refrigeration capacity. (After [52.21])

stimulated *J. G. Bednorz* and *K. A. Müller* to search for higher critical temperatures in oxides. Their discovery of high-temperature superconductivity in the cuprate $(\text{La, Ba})_2\text{CuO}_4$ ($T_c \approx 30$ K) opened up a new field of research [52.12]. Within less than a year a critical temperature well above the boiling point of liquid nitrogen could be achieved in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ [52.20]. So far, the highest transition temperature of 135 K at ambient pressure has been found in $\text{HgBa}_2\text{Ca}_2\text{Cu}_3\text{O}_{8+\delta}$ [52.13, 14].

Finally, it should be mentioned that the phenomenon of superconductivity has also been observed in organic materials [52.22–24] and fullerenes [52.25–27]. Remarkably high transition temperatures of 19, 33 and 40 K

have been found in the fullerenes K_3C_{60} , $\text{RbCs}_2\text{C}_{60}$ and Cs_3C_{60} respectively.

An important advantage of high-temperature superconductors is the possibility of operating devices at temperatures well above 20 K, leading to a considerable reduction in the required refrigerator input power. In contrast to power applications, the use of liquid nitrogen as a coolant is of less importance for superconducting electronics, which are typically cooled by closed cycle coolers. The economics of closed cycle coolers depends strongly on the operating temperature. Considering an ideal, reversible cooling cycle, the input power necessary to remove a heat load dQ/dt is

$$P_{\text{in}} = \frac{1}{\eta_C} \frac{dQ}{dt}, \quad (52.1)$$

where

$$\eta_C = \frac{T_{\text{op}}}{300 \text{ K} - T_{\text{op}}} \quad (52.2)$$

is the Carnot efficiency. To remove a heat input of 1 W an ideal, reversible refrigerator consumes at room temperature a power of 70 W for $T_{\text{op}} = 4.2$ K, whereas this power is only 9 and 2.9 W for operation at 30 and 77 K, respectively. The efficiency η_r of real refrigerators is much smaller than the Carnot efficiency and depends strongly on the refrigerator size, as illustrated in Fig. 52.3. For small refrigerators able to remove a power of 1 W, the refrigerator efficiency is only a few percent of the Carnot efficiency. On the other hand, large refrigerators with a cooling capacity well above 10 kW may reach 30% of the Carnot efficiency. The ratio η_r/η_C has been found to depend only weakly on the operating temperature.

52.1 The Superconducting State

52.1.1 Characteristic Properties of Superconductors

The discussion of the physical properties of high-temperature superconductors has to be based on the knowledge of the behaviour of conventional, metallic superconductors. In this section the superconducting state is therefore described in more detail starting, from the properties of simple, metallic superconductors.

Below a certain temperature T_c the electrical resistance of a superconducting material vanishes. The critical temperature T_c is characteristic of the superconductor in question. In pure metals the zero resistance

state can be reached within a few mK (Fig. 52.1). In the chemically complex cuprate, high- T_c superconductors the transition to the superconducting state is less sharp than in metallic low- T_c superconductors. The transition width ΔT_c for single-phase cuprate superconductors is typically 1 K. Thus the critical temperature depends slightly on the criterion used to define T_c . Several criteria are illustrated in Fig. 52.4.

At the transition temperature a resistance drop of several orders of magnitude is typically observed. However, it is in principle not possible to prove experimentally that the resistance in the superconducting state is in fact zero. The most efficient way to determine

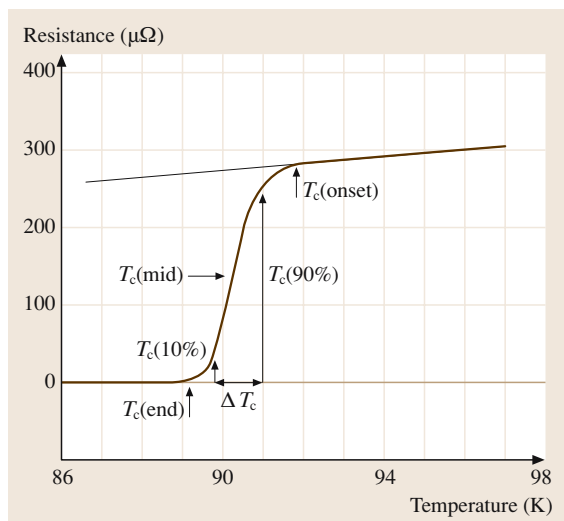


Fig. 52.4 Resistance versus temperature for a $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8/\text{Ag}$ multi-core wire. The transition temperatures resulting from different definitions of T_c are indicated. The width of the transition $\Delta T_c = T_c(90\%) - T_c(10\%)$ is ≈ 1.2 K

an upper limit of the resistance is to detect the decay of the magnetic fields generated by currents induced in a superconducting loop. Upper resistivity limits between 2×10^{-18} [52.28] and $7 \times 10^{-23} \Omega \text{ cm}$ [52.29] have been reported for the high- T_c superconductor $\text{YBa}_2\text{Cu}_3\text{O}_7$, while a value of $3.6 \times 10^{-23} \Omega \text{ cm}$ has been found for type I low- T_c superconductors [52.30]. These upper resistivity limits are orders of magnitude smaller than the resistivity of $10^{-10} \Omega \text{ cm}$ at 4.2 K achievable in annealed, very pure metals. It is therefore justified to assume zero resistance below T_c for all practical purposes.

The second outstanding feature of the superconducting state is perfect diamagnetism. To understand the magnetic behaviour of a superconductor, the two different situations illustrated in Fig. 52.5 have to be considered. In the first case the superconductor is cooled below T_c without an applied magnetic field (zero field-cooled). In the second case the superconductor is cooled below T_c while a magnetic field is applied (field-cooled). In both cases the magnetic flux is excluded from the interior of the superconductor. Screening currents flowing in the surface layer of the superconductor generate a flux opposite to that of the applied magnetic field. The magnetic flux density is therefore zero everywhere inside the superconductor. On the other hand, outside the superconducting sphere the magnetic field is enhanced because of the superposition of the flux of the

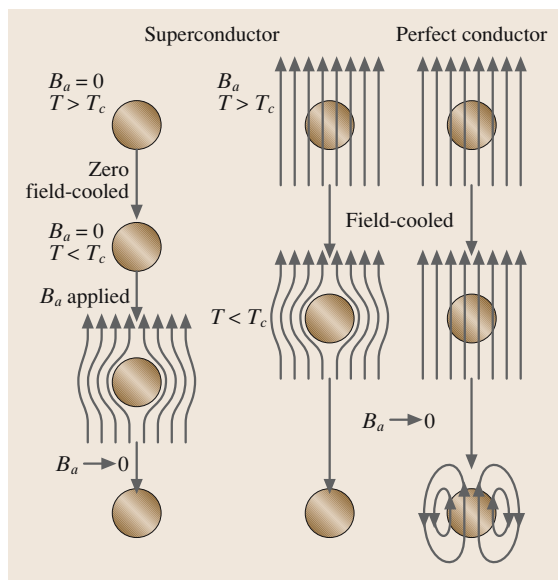


Fig. 52.5 The magnetic flux is excluded from the interior of a superconductor without field-cooling (*left*) as well as with it (*centre*). In contrast to this behaviour, a magnetic flux would exist in the interior of a field-cooled perfect conductor (*right*) ([52.31] p. 20)

applied magnetic field and that of the screening currents. In both situations the superconductor is unmagnetised when the magnetic field is removed. The behaviour of a zero field-cooled superconductor may be considered simply as a screening effect caused by perfect conductivity. In contrast to screening, the expulsion of the magnetic flux from a field-cooled superconductor (the Meissner effect) cannot be explained by perfect conductivity. The different behaviour of a field-cooled perfect conductor is also shown in Fig. 52.5. The values of the relative magnetic permeability for non-ferromagnetic metals is close to unity. Thus the magnetic flux inside the metal is nearly the same as that of the applied magnetic field. Because $dB/dt = 0$ no screening currents are induced, and hence the magnetic flux is not removed from the interior of the perfect conductor at low temperatures. Finally, the removing of the magnetic field ($dB/dt \neq 0$) leads to magnetisation of the perfect conductor.

The superconducting state can be destroyed by a sufficiently large applied magnetic field. The transition from the superconducting to the normal state occurs at a certain critical magnetic field B_c , which is a characteristic property of the superconductor considered. The dependence of the critical field on the temperature is

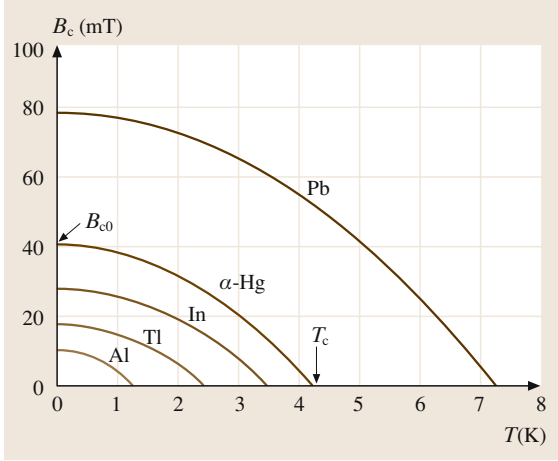


Fig. 52.6 Critical field versus temperature curves for selected metallic superconductors. The B_{c0} values vary from ≈ 10 to ≈ 80 mT ([52.31] p. 24)

typically well described by

$$B_c(T) = B_{c0} \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \quad (52.3)$$

where T_c is the critical temperature and B_{c0} the critical field at zero temperature. The critical field as a function of temperature is shown in Fig. 52.6 for selected metallic superconductors. The $B_c(T)$ curves separate the superconducting from the normal state. The Meissner effect ensures that the properties at any point in the superconducting state are independent of the order in which the final conditions of applied magnetic field and temperature are reached.

52.1.2 Superconductor Electrodynamics

The electrodynamics of superconductors can be described by the two London equations. As a consequence of the infinite conductivity in the superconducting state, Ohm's law must be replaced by the first London equation

$$\frac{d\mathbf{j}}{dt} = \frac{1}{\mu_0 \lambda_L^2} \mathbf{E}, \quad (52.4)$$

where \mathbf{j} is the current density, \mathbf{E} the electric field, μ_0 is the permeability of free space and λ_L a constant. The electric field in the superconductor is zero for constant current density. The current density and the magnetic field are related by the second London equation.

$$\text{curl} \mathbf{j} = -\frac{1}{\mu_0 \lambda_L^2} \mathbf{B}. \quad (52.5)$$

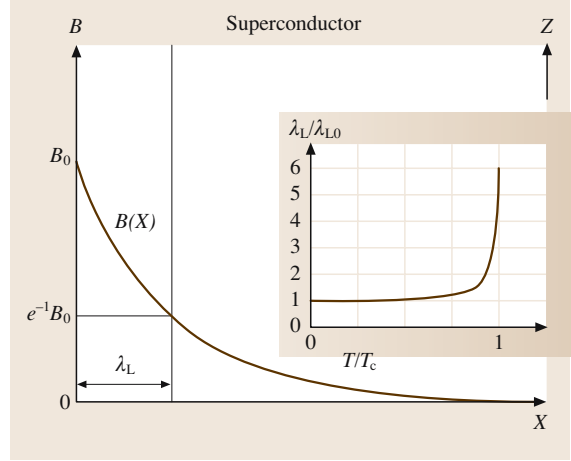


Fig. 52.7 Exponential decay of an applied magnetic field in the surface layer of a semi-infinite superconductor. The normalised penetration depth λ_L/λ_{L0} versus reduced temperature is shown in the inset

Using the Maxwell equation $\text{curl} \mathbf{B} = \mu_0 \mathbf{j}$ the following differential equation results for the magnetic field

$$\nabla^2 \mathbf{B} = \frac{1}{\lambda_L^2} \mathbf{B}. \quad (52.6)$$

For a semi-infinite superconductor with the magnetic field applied along the z -direction, an exponential decay of the magnetic field $B(x) = B_0 \exp(-x/\lambda_L)$ results, as illustrated in Fig. 52.7. The London penetration depth λ_L is therefore the characteristic length for the decay of the magnetic field in the surface layer of the superconductor. Typical values of the penetration depth are 10^{-5} – 10^{-6} cm at $T = 0$. The penetration depth is closely related to the number density of the Cooper pairs (Sect. 52.1.3)

$$\lambda_L = \sqrt{\frac{m_C}{\mu_0 n_C q_C^2}}, \quad (52.7)$$

where n_C , q_C and m_C are the number density, the charge and the mass of the Cooper pairs. Because of the decreasing Cooper pair density, the penetration depth increases rapidly for temperatures close to T_c (see inset of Fig. 52.7). The temperature dependence is given by

$$\lambda_L(T) = \lambda_{L0} \left[1 - \left(\frac{T}{T_c} \right)^4 \right]^{-1/2}, \quad (52.8)$$

where λ_{L0} is the penetration depth at $T = 0$.

52.1.3 Superconductivity: A Macroscopic Quantum Phenomenon

Bardeen–Cooper–Schrieffer (BCS) Theory

Next, the microscopic causes of superconductivity in conventional superconductors are briefly described. *J. Bardeen, L. N. Cooper, and J. R. Schrieffer* proposed a microscopic theory of superconductivity [52.4, 5], which is now well known as the BCS theory. At the moment it seems to be unlikely that the mechanism of superconductivity in the cuprates, which is still not yet understood, can be explained by the BCS theory. In this chapter only some fundamental ideas and results of the BCS theory are presented.

Independently of each other, two researchers [52.32, 33] found theoretically that lattice vibrations (phonons) can provide an attractive interaction between two electrons. This can be understood as a consequence of the polarisation of the lattice by the electrons. The concentration of positive charges produced by the first electron reduces the potential energy of the second electron. Because of this attractive interaction, electrons with opposite momenta and spins form electron (Cooper) pairs. The Cooper pairs, characterised by zero spin and momentum, have to be considered as new particles. Because their spin is zero, they do not obey Pauli's exclusion principle. The probability that a state is occupied increases with the number of particles populating the same quantum state. Therefore, all Cooper pairs occupy the same quantum state with a single value of the momentum in the presence of an applied electric field. A change of momentum results only for scattering processes that break up some Cooper pairs. Resistance occurs as soon as the energy taken from the electric field exceeds the binding energy of the Cooper pairs. This behaviour explains the existence of a critical current.

The electron–phonon interaction leads to an energy gap 2Δ in the density of states for the single electrons. The width of the energy gap at zero temperature is

$$2\Delta(0) = 3.5k_B T_c, \quad (52.9)$$

where k_B is the Boltzmann constant and T_c the critical temperature. For superconducting elements the measured ratios of $2\Delta(0)/k_B T_c$ are typically 3.2–4.6 [52.34], which is in reasonably good agreement with the value of 3.5 predicted by the BCS theory. The single-electron density of states close to the Fermi energy is

$$D_s(E - E_F) = D_n(E_F) \frac{|E - E_F|}{\sqrt{(E - E_F)^2 - \Delta^2}}, \quad (52.10)$$

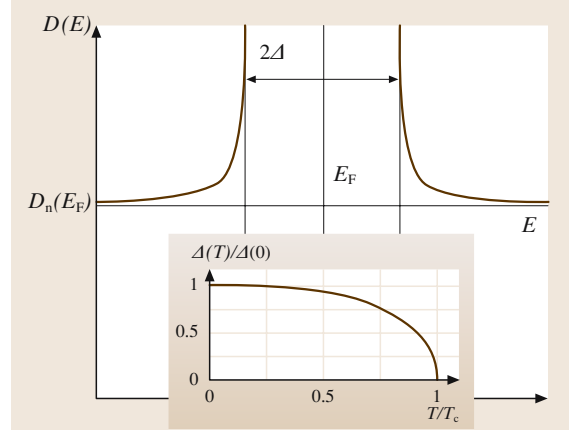


Fig. 52.8 Energy gap and density of states for unpaired electrons in the superconducting state in the vicinity of the Fermi energy E_F . The temperature dependence of the energy gap predicted by the BCS theory is shown in the inset

where Δ is the half-width of the energy gap and $|E - E_F| > \Delta$. The variation of the normal state density of states $D_n(E)$ can be neglected because $E_F \gg 2\Delta$. The single-electron density of states shown in Fig. 52.8 reflects that the energy states present in the normal state within the energy gap are shifted to the edges. The inset of Fig. 52.8 shows the temperature dependence of the energy gap predicted by the BCS theory. The critical temperature is [52.5]

$$T_c = 1.14 \frac{\hbar\omega_D}{k_B} \exp\left(-\frac{1}{D_n(E_F)V_p}\right), \quad (52.11)$$

where the electron–phonon interaction is described by the constant V_p . The Debye frequency ω_D is the cut-off frequency of the phonon spectrum in a solid. Considering the atoms in a solid as harmonic oscillators of frequency $\omega = (k_e/M)^{1/2}$, where k_e is the elastic constant resulting from the binding forces, both ω_D and T_c are expected to be proportional to $M^{-1/2}$ for different isotopes of the same element. In many metallic superconductors this isotope effect has been observed [52.35–38].

All Cooper pairs populate the same quantum state, resulting in a collective wave function

$$\psi = n_C^{1/2} \exp[i\varphi(\mathbf{r})], \quad (52.12)$$

where $n_C = \psi^*\psi$ (ψ^* : conjugate complex of ψ) is their concentration. Because the phase coherence of the Cooper pairs is effective at large distances, the phase

$\varphi(\mathbf{r})$ is a macroscopically observable quantity. As a consequence, the magnetic flux in a superconducting hollow cylinder of cross-sectional area F is quantised

$$\phi = \int_F \mathbf{B} \cdot d\mathbf{F} = \frac{h}{2e} N_\phi = \phi_0 N_\phi. \quad (52.13)$$

In the above equation N_ϕ is an integer, e the electron charge and $\phi_0 = 2.068 \times 10^{-15} \text{ V s}$ the flux quantum. The magnetic flux generated by the screening currents flowing in the surface layer of the superconductor is typically small, and has therefore been neglected.

Tunnelling Effects

First, single-electron tunnelling in a normal–insulator–superconductor (N-I-S) junction will be considered. The two conductors are separated by an insulating oxide layer typically less than 2 nm thick. Tunnelling can take place only when empty target energy states are available, otherwise the process is forbidden by the Pauli principle (Fig. 52.9). For a voltage of $U = \Delta/e$ ($-e > 0$) the empty energy states of the superconductor above the energy gap reach the same level as the Fermi energy of the normal conductor, and strong tunnelling sets in. On the other hand for a voltage $U = -\Delta/e$, pair breaking becomes possible. One of the electrons tunnels into the normal conductor and gains an energy Δ available for the exci-

tation of the second electron to an empty state above the energy gap. The resulting N-I-S current–voltage characteristic and the Bose condensation representation of the tunnelling process are shown in Fig. 52.9.

In the case of two identical superconductors (S-I-S junction), strong single-electron tunnelling sets in at an applied voltage $|U| \geq |2\Delta/e|$. This voltage is high enough to allow the empty energy states above the gap of the first superconductor, and the filled energy states below the gap of the second superconductor, to overlap; pair breaking becomes possible. Figure 52.10 shows the semiconductor representation of the tunnelling process and the resulting S-I-S current–voltage characteristic. The weak tunnelling currents for $|U| < |2\Delta/e|$ are a consequence of a few quasi-particles being thermally excited above the energy gap for $T > 0$ (Fig. 52.10).

So far, only single-electron tunnelling has been considered. In 1962 *B. D. Josephson* predicted that Cooper pair tunnelling through a very thin insulating layer is possible [52.10]. Even for zero voltage across the insulating layer of an S-I-S junction, Cooper pairs can tunnel through the barrier. This process has to be considered as

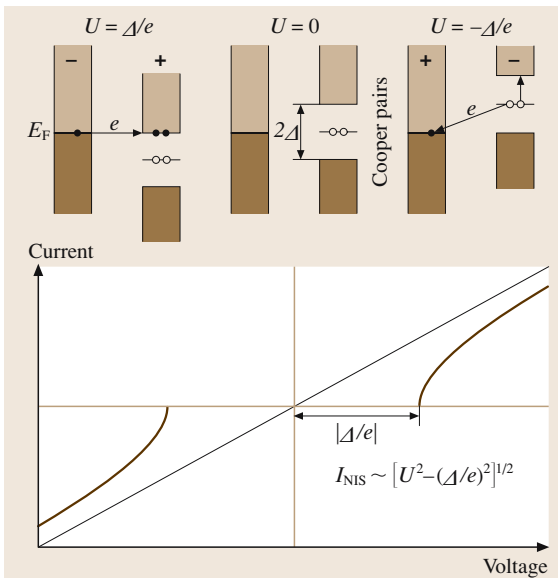


Fig. 52.9 Bose condensation representation of N-I-S tunnelling (top) and the corresponding current voltage characteristic at $T = 0$

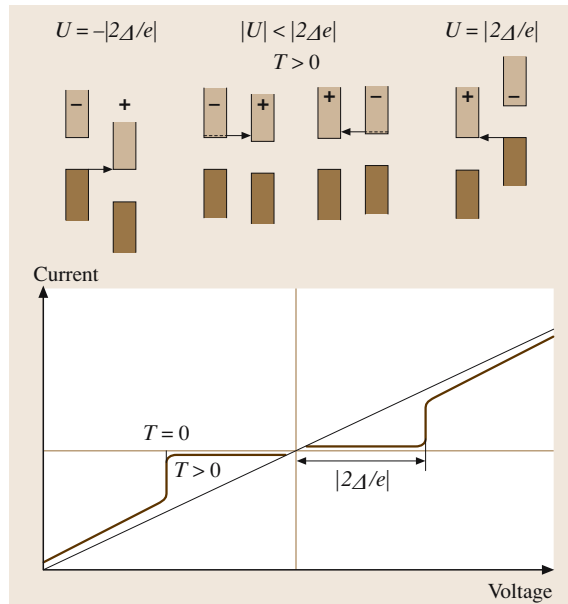


Fig. 52.10 Current voltage characteristic of an S-I-S tunnel junction (bottom). Strong tunnelling sets in for $U = |2\Delta/e|$. The semiconductor representation of the process (top) shows that a few electrons (quasi particles) are thermally excited into states above the gap at $T > 0$. These electrons are responsible for the weak tunnelling current at low voltages ([52.31] p. 42)

the tunnelling of the wave function collectively describing the Cooper pairs. The DC current through the contact is

$$I = I_c \sin \delta, \quad (52.14)$$

where I_c is the critical current of the Josephson junction. The Josephson DC current through the contact is between $-I_c$ and I_c , depending on the phase difference δ of the two Cooper pair wave functions. This phase difference is time independent for zero voltage across the insulating layer. Applying a DC voltage U across the insulating layer of the Josephson contact causes a time-dependent phase difference between the two Cooper pair wave functions. As a consequence there flows an AC current

$$I_{AC} = I_c \sin [2\pi\nu_J t + \delta(0)], \quad (52.15)$$

where

$$\nu_J = \frac{2|e|U}{h} = \frac{U}{\phi_0}. \quad (52.16)$$

The AC Josephson current I_{AC} oscillates with a frequency of 483.6 MHz/ μ V. The tunnelling of the Cooper pair through the barrier is connected to the emission or absorption of a photon with energy $E = h\nu_J$. The AC Josephson current flows in addition to the current resulting from single-electron tunnelling. A more detailed description of the AC Josephson effect can be found in [52.34].

Finally, macroscopic quantum interference effects in Josephson junctions will be briefly considered. These interference effects are closely connected to the phase shifts of the Cooper pair wave functions caused by a magnetic field applied transversely to the current direction. The applied magnetic field gives rise to a spatial variation of the phase difference of the two Cooper wave functions within the contact area. The resulting Josephson DC current for a rectangular barrier is

$$I = I_c \sin \delta_0 \frac{\sin \pi\phi/\phi_0}{\pi\phi/\phi_0}, \quad (52.17)$$

where I_c is the critical current of the Josephson junction, ϕ the total flux enclosed by the barrier, and ϕ_0 the flux quantum. The phase shift δ_0 is unknown. The maximum Josephson DC current

$$I_{\max} = I_c \left| \frac{\sin \pi\phi/\phi_0}{\pi\phi/\phi_0} \right| \quad (52.18)$$

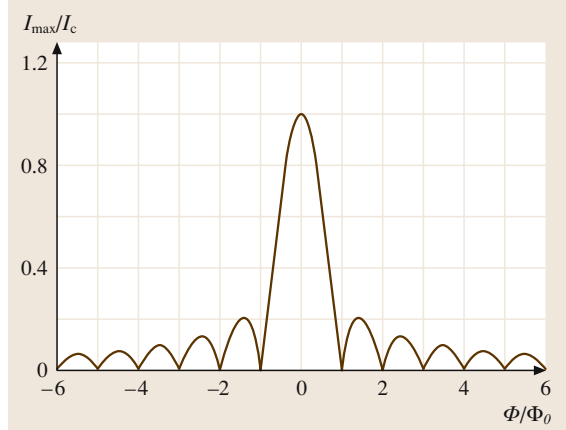


Fig. 52.11 Josephson–Fraunhofer diffraction pattern for the normalised maximum Josephson DC current I_{\max}/I_c . Minima of the current occur for Φ/Φ_0 values of 1, 2, 3 etc.. (After [52.31] p. 49)

is reached for $\sin \varphi_0 = 1$, corresponding to a phase difference of $\pi/2$. In Fig. 52.11, the resulting interference pattern for the Josephson DC current is shown. The Josephson–Fraunhofer interference of the two Cooper pair wave functions indicates that superconductivity is a macroscopic quantum phenomenon (for more details see [52.31]).

52.1.4 Type II Superconductors

A quantitative description of the behaviour of type II superconductors is provided by the Ginzburg–Landau theory (see for example [52.34]). In the London theory the Cooper pair density is assumed to be constant. The only characteristic length scale is the penetration depth λ_L describing the decay of a magnetic field inside a superconductor. Taking into consideration a variation of the Cooper pair density leads to a second characteristic length scale, namely the coherence length ξ , the shortest distance within which a considerable change of the Cooper pair density can occur. At the interface of a normal and a superconductor the Cooper pair density increases continuously from zero to the value n_C deep inside the superconductor. The ratio of the coherence length and the penetration depth is the Ginzburg–Landau parameter κ

$$\kappa = \frac{\lambda_L}{\xi}. \quad (52.19)$$

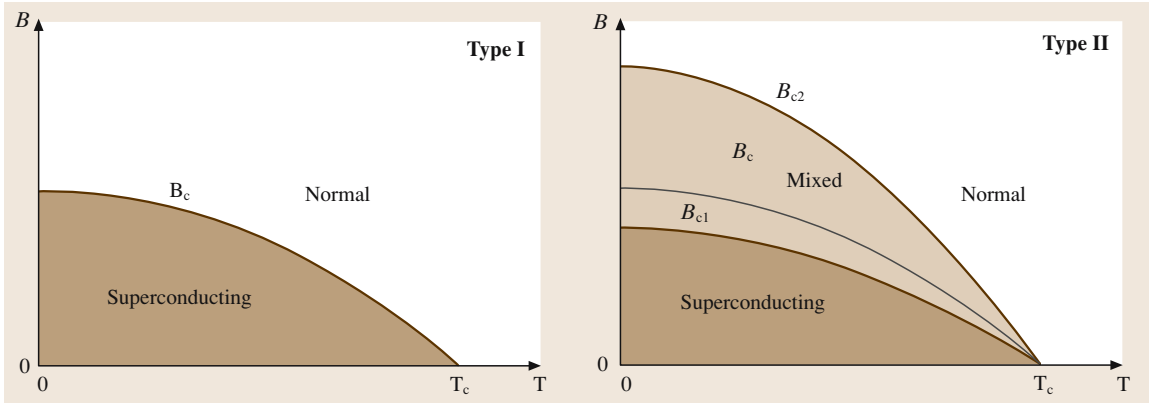


Fig. 52.12 Comparison of the phase diagrams for type I and II superconductors. (After [52.31] p. 74)

Depending on whether the parameter κ is smaller or larger than $1/\sqrt{2}$, the superconductor in question is of type I or II, respectively. Values of the coherence length and the Ginzburg–Landau parameter of selected metallic superconductors are listed in Table 52.2. The cuprate high- T_c superconductors are extreme type II superconductors with κ values of ≈ 100 .

In Fig. 52.12, the phase diagrams of both types of superconductors are compared. In type I superconductors there exist two different phases, namely the superconducting and the normal, which cannot coexist. On the other hand, three different phases can exist in type II superconductors. Below the lower critical field $B_{c1}(T)$ the type II superconductor is in the superconducting phase characterised by perfect diamagnetism, where it behaves like a type I superconductor. Above the upper critical field $B_{c2}(T)$ the material is in the normal state. In the mixed state between the lower and upper critical fields superconducting and normal regions can coexist in a type II superconductor. The mixed state shown in Fig. 52.13 is characterised by an array of normal cores within the superconductor, which allow magnetic flux to enter into the superconductor. The supercurrents encircling the normal cores (vortices) cause a repulsive vortex–vortex interaction.

In the mixed state of a type II superconductor the transport currents can flow throughout the conductor. In the presence of a magnetic field perpendicular to the current direction a Lorentz force $F_L = j\phi_0$ acts

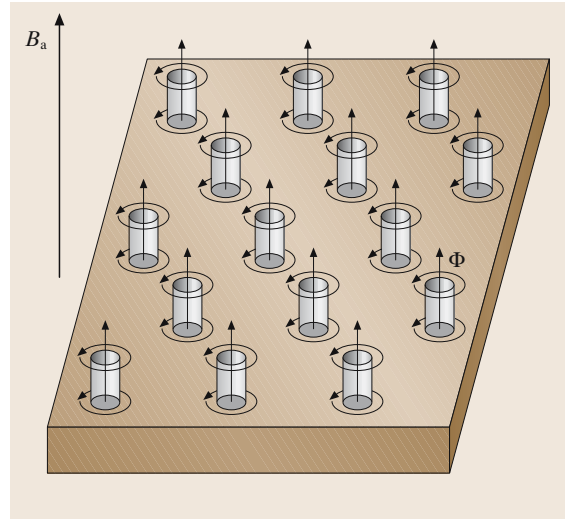


Fig. 52.13 Type II superconductor in the mixed state. Each vortex contains just a single fluxoid $\Phi_0 = h/2|e|$. (After [52.31] p. 74)

on the vortices, which leads to a motion of the vortices perpendicular to the current and the field direction. The variation of the magnetic flux leads to a flux flow resistance, and hence the critical current would be zero. However, in a real type II superconductor defects are always present, which act as pinning centres. As long as the Lorentz force is smaller than the pinning force, flux motion is suppressed, resulting in zero resistance.

52.2 Cuprate High- T_c Superconductors: An Overview

52.2.1 Major Families of Cuprate Superconductors

At present around 100 cuprate superconductors are known. The major families of cuprate superconductors are listed in Table 52.3. Common features of the cuprate superconductors are layered crystal structures. In the cuprates, conducting copper oxide blocks and insulating charge-carrier reservoirs alternate with each other along the crystallographic c direction. As a consequence of the layered structures the physical properties of the cuprates are highly anisotropic. The electrical resistivity along the crystallographic a - and b -directions within the conductive CuO_2 planes is much lower than along the c -direction, where the current has to cross the insulating charge-carrier reservoirs. Figure 52.14 shows as an example the layering scheme of $\text{Tl-}22(n-1)n$ superconductors. The three thallium-based compounds are distinguished by the different number n of CuO_2 planes in the copper oxide blocks. In the compounds with $n \geq 2$, adjacent CuO_2 planes are separated by Ca layers. In the $\text{Bi-}22(n-1)n$ counterparts Tl and Ba are replaced by Bi and Sr, respectively. In both families of cuprate superconductors the copper oxide blocks are embedded between TlO or BiO double layers. An analogous layering scheme results for the $\text{Tl-}12(n-1)n$ compounds. However, the copper oxide blocks of these compounds are separated only by single TlO layers. The highest critical temperature is typically reached in the compounds with three CuO_2 layers ($n = 3$) in the copper oxide blocks.

Table 52.2 Coherence length ξ_c and Ginzburg–Landau parameter κ for selected superconductors at $T = 0$ ([52.31] p. 70)

Material	$T_c(\text{K})$	$\xi(\text{nm})$	$\kappa = \lambda_L/\xi$	Reference
Al	1.18	550	0.03	[52.34]
In	3.41	360	0.11	[52.34]
Cd	0.52	760	0.14	[52.34]
Sn	3.72	180	0.23	[52.34]
Ta	4.5	93	0.38	[52.39]
Pb	7.2	82	0.48	[52.34]
Nb	9.25	39	1.28	[52.34]
NbTi	9.5	4	75	[52.40]
Nb ₃ Sn	18	3	21.7	[52.40]
Nb ₃ Ge	23.2	3	30	[52.40]
Rb ₃ C ₆₀	29.6	2	123.5	[52.41]
YBa ₂ Cu ₃ O _{7-δ}	93	2	95	[52.34]

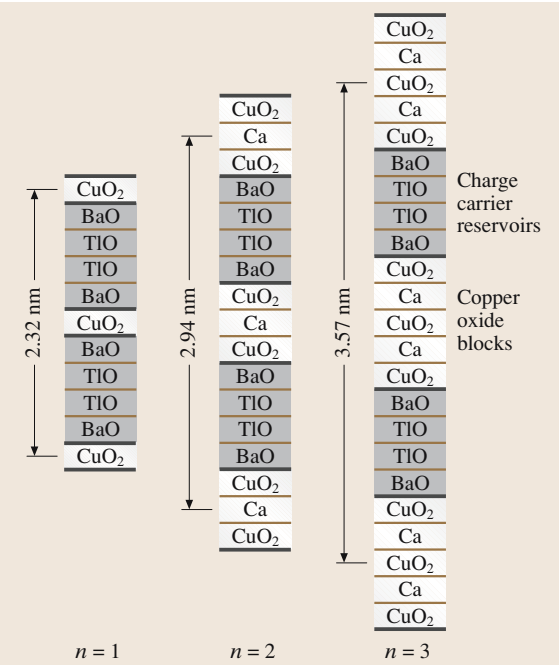


Fig. 52.14 Layering scheme of $\text{Tl}_2\text{Ba}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_{2n+4}$ superconductors (After [52.31] p. 81)

The physical properties of the cuprate superconductors depend strongly on their chemical composition and the resulting carrier concentration. As in semiconductors, the carrier concentration in the cuprates can be changed by doping. For example let us consider the superconductor $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. The possible oxidation states of copper are +1, +2 and +3. In the insulating parent compound La_2CuO_4 the oxidation state of copper is +2. The valence of La is +3 and that of Sr +2. Substitution of Sr^{2+} for La^{3+} leads to the formation of Cu^{3+} or O^- defects (holes), which are expected to be mobile. Thus, an increasing number of holes is created in the CuO_2 planes with increasing Sr concentration x .

The cuprates are complex non-stoichiometric oxides. Heat treatments in reducing or oxidising atmosphere therefore lead to considerable changes in the oxygen content. Incorporation of excess oxygen is a further way to enhance the hole concentration in the CuO_2 planes of the cuprates.

Finally, it should be noted that, in addition to the major families of cuprate superconductors listed in Table 52.3, there exists a large number of high-

temperature superconductors which can be synthesised only under high pressure (see for example [52.31]).

52.2.2 Generic Phase Diagram of Cuprate Superconductors

It is beyond the scope of this article to discuss all the new physical phenomena related to cuprate superconductors. To give an impression of the complexity of

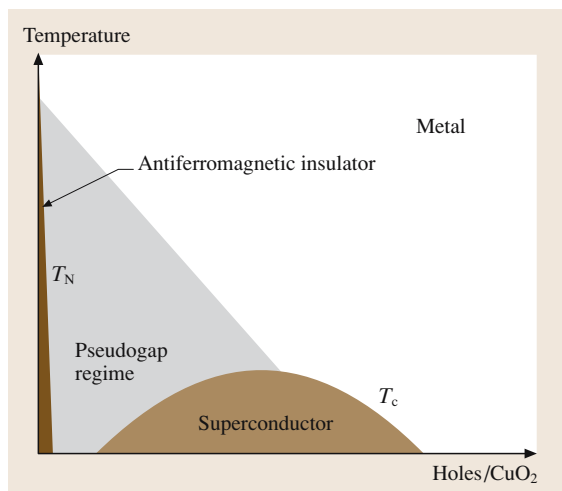


Fig. 52.15 Generic phase diagram for cuprate superconductors indicating that these materials are close to a metal–insulator transition. The physical properties of the cuprates vary considerably with increasing hole concentration

these materials the generic phase diagram is briefly described. The one shown in Fig. 52.15 indicates that the cuprate superconductors are close to a metal–insulator transition. Their parent compounds are antiferromagnetic insulators, of which La_2CuO_4 and $\text{YBa}_2\text{Cu}_3\text{O}_6$ are examples. The Néel temperature T_N , at which magnetic ordering sets in, reaches values of 250–400 K in the parent compounds. The Néel temperature decreases rapidly with increasing hole concentration. A transition from an insulating to a superconducting phase is observed for further enhanced doping levels. The critical temperature increases with increasing number of holes and reaches a maximum at ≈ 0.18 holes per CuO_2 unit before decreasing again. At doping levels well above 0.3 holes/ CuO_2 the material behaves as a non-superconducting metal.

In the last few years it could be clarified that the energy gap in cuprate superconductors is of d-wave symmetry, whereas it is of s-wave symmetry in conventional superconductors. The two different symmetries are illustrated in Fig. 52.16. For s-wave symmetry the superconducting energy gap is constant irrespective of the direction of the wave vector k . In the case of d-wave symmetry there exist positions on the Fermi surface for which the energy gap is zero. At these positions the sign of the order parameter changes from + to – or vice versa. As a consequence a finite tunnel conductivity [$dI/dV \propto D(E)$] is observed within the energy gap (see Fig. 52.16).

Next, let us consider the temperature dependence of the energy gap. The tunnelling conductance of underdoped Bi-2212 shown in Fig. 52.17 indicates that

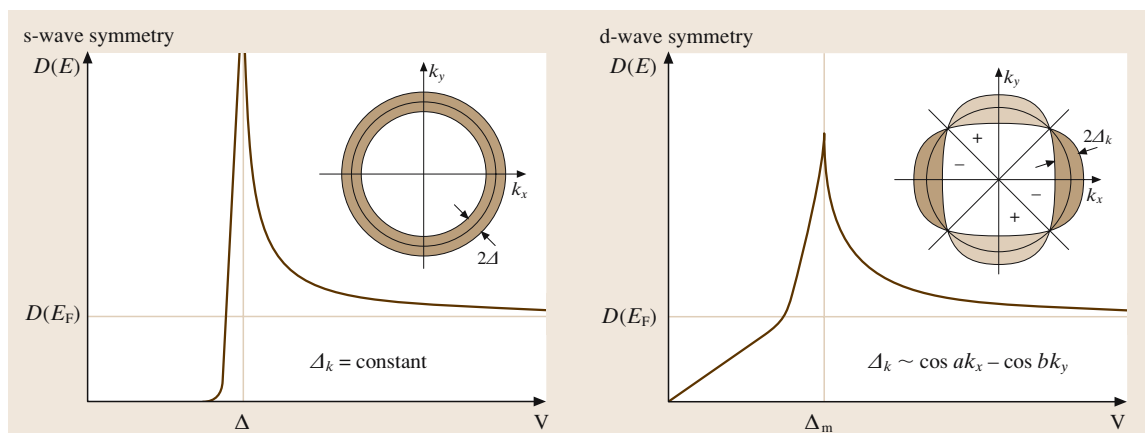


Fig. 52.16 The two different symmetries of the energy gap and the corresponding density of states. For s-wave symmetry (left) the width of the energy gap is independent of the wave vector k . In the case of the d-wave symmetry (right) the energy gap depends on the wave vector. At four positions on the Fermi surface the energy gap is zero (After [52.42])

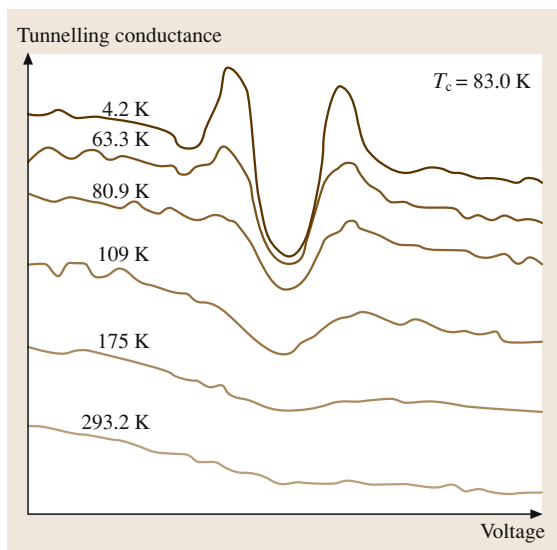


Fig. 52.17 Schematic illustration of the tunnelling conductance of underdoped Bi-2212 providing the density of states in the vicinity of the energy gap. For clarity the spectra are offset vertically. A gap-like feature (pseudogap) can be observed even at a temperature as high as 175 K, well above T_c (After [52.43])

the energy gap of cuprate superconductors is independent of temperature. This behaviour is in contrast to the predictions of the BCS theory. In underdoped cuprate superconductors the energy gap has been found to increase for decreasing number of charge carriers in spite of a reduced critical temperature. On the other hand, both the energy gap and the critical temperature of overdoped cuprates decrease with increasing carrier concentration. An outstanding feature of cuprate superconductors is the existence of an energy gap even at temperatures well above T_c . This pseudogap is found in the region of the phase diagram where a transition from an antiferromagnetic insulator to a superconductor occurs.

New theories have predicted that in underdoped cuprate superconductors, spin and charge are separated, resulting in antiferromagnetic domains and charge stripes located at the anti-phase boundaries [52.44]. Experimental evidence for this scenario has been found from neutron scattering [52.45]. A potential explanation for the existence of the pseudogap is the formation of Cooper pairs at temperatures well above T_c . However, because of the small Cooper pair density, they are not macroscopically correlated. As a consequence the material is not able to carry a supercurrent above T_c . At present it is not clear if the interplay of anti-

ferromagnetism and charge stripes is responsible for high-temperature superconductivity.

52.2.3 Crystal Structures

Next, let us consider the crystal structures of the cuprates in more detail. For $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (Y-123) an orthorhombic structure with lattice parameters $a = 0.38 \text{ nm}$, $b = 0.39 \text{ nm}$ and $c = 1.17 \text{ nm}$ has been deduced from X-ray powder diffraction data [52.46, 47]. Figure 52.18 shows the orthorhombic unit cell of Y-123. The unit cell of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ contains double CuO_2 layers with embedded yttrium ions (Fig. 52.18). Slightly different z positions of the Cu and O atoms lead to buckling in the CuO_2 planes. The oxygen atoms occupy perovskite-like anion positions halfway between the copper atoms. These square-coordinated CuO_2 layers are a characteristic feature of all cuprate superconductors. Because metallic conductivity has been found in these layers, they are also called conduction planes. Along the crystallographic c -axis perpendicular to these planes, insulating or semiconducting behaviour

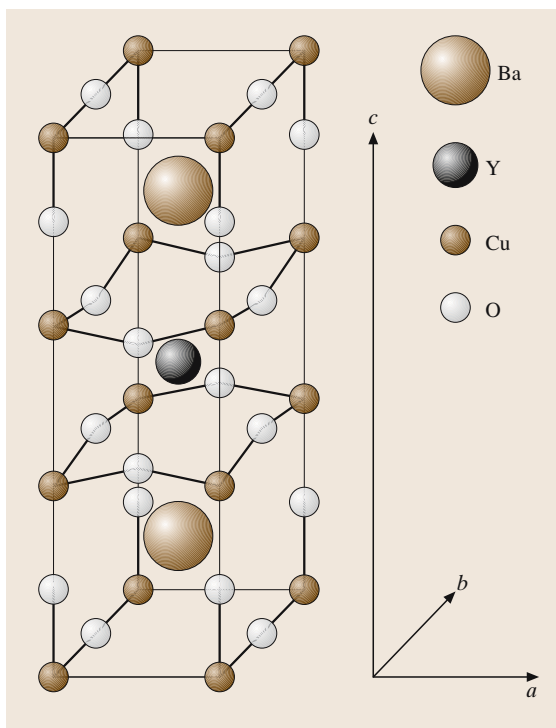


Fig. 52.18 Orthorhombic unit cell of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. A special feature of this cuprate are CuO chains in the b -direction (After [52.31] p. 100)

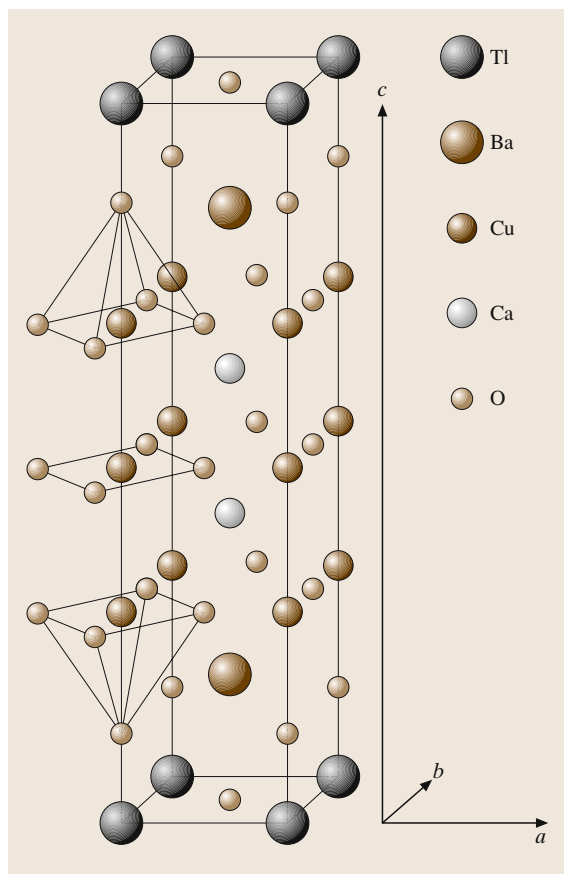


Fig. 52.19 Primitive tetragonal unit cell of $\text{TlBa}_2\text{Ca}_2\text{Cu}_3\text{O}_9$ (After [52.31] p. 108)

is frequently observed. Neighbouring CuO_2 blocks are separated by two BaO layers. The CuO chains between the two BaO layers in the b -direction are a special feature of the Y-123 compound.

The crystal structures of $\text{Tl-}12(n-1)n$ and the analogous $\text{Hg-}12(n-1)n$ superconductors are primitive tetragonal. Because of the common CuO_2 planes, the a and b lattice parameters of cuprate superconductors are generally close to 0.39 nm. In both cuprate superconductor families the c lattice parameter is given approximately by $c \cong (0.63 + 0.32n)\text{nm}$. As an example of these crystal structures the primitive tetragonal unit cell of $\text{TlBa}_2\text{Ca}_2\text{Cu}_3\text{O}_9$ is shown in Fig. 52.19. The copper oxide blocks of the Tl-1223 compound are formed of three CuO_2 layers. The CuO_2 layers within the copper oxide blocks are separated by Ca ions. In contrast to the $\text{Tl-}22(n-1)n$ compounds (see Fig. 52.14) adjacent CuO_2 blocks are separated by only

Table 52.3 Major families of superconducting cuprates ([52.31] p. 79)

Family	Symbol	Maximum T_c (K)
$(\text{La}_{1-x}\text{M}_x)_2\text{CuO}_4^{\text{a}}$	La-214	39
$\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}^{\text{b}}$	Y-123	92
$\text{YBa}_2\text{Cu}_4\text{O}_8$	Y-124	80
$\text{Bi}_2\text{Sr}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_{2n+4}$	Bi-22($n-1$) n	122
$\text{Tl}_2\text{Ba}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_{2n+4}$	Tl-22($n-1$) n	128
$\text{TiM}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_{2n+3}^{\text{a}}$	Tl-12($n-1$) n	122
$\text{HgBa}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_{2n+2}$	Hg-12($n-1$) n	135

^aM = Sr or Ba, ^bY can be replaced by rare-earth elements

a single TlO layer. Like the TlO double layers of the $22(n-1)n$ compounds, the single TlO layers of the $12(n-1)n$ compounds are embedded between two BaO layers. The reduced distance of neighbouring copper oxide blocks leads to an enhanced coupling of the conduction planes. This improved coupling is reflected in less-anisotropic superconducting properties (Sect. 52.3). Defects in the $\text{BaO}/\text{TlO}/\text{BaO}$ layers forming the charge-carrier reservoirs provide holes for the doping of the CuO_2 planes. In the $\text{Tl-}12(n-1)n$ compounds the BaO layers can be replaced by SrO layers. The substitution of Tl by Hg leads to the structurally analogous $\text{Hg-}12(n-1)n$ counterparts. The lattice parameters of selected cuprate superconductors are listed in Table 52.4.

52.2.4 Critical Temperatures

Next, some of the factors determining the critical temperature of cuprate superconductors will be discussed. As shown in the generic phase diagram for cuprate superconductors (Fig. 52.15), the critical temperatures of these materials are closely connected to the hole concentration p in the CuO_2 planes, which is determined by the formal valence $2 + p$ of the copper atoms in the conduction planes. Generally, an inverse parabolic dependence of the critical temperature on the number of holes per CuO_2 has been found. The maximum T_c is typically reached for a hole concentration p of 0.15–0.2 holes per CuO_2 . For hole concentrations above 0.2 holes per CuO_2 the critical temperature decreases with increasing carrier concentration, leading to a non-superconducting metal for $p > 0.3$ holes per CuO_2 . As an example of this behaviour the critical temperature of $\text{HgBa}_2\text{CuO}_{4+\delta}$ as function of the hole concentration is shown in Fig. 52.20.

Table 52.4 Lattice parameters of selected cuprate superconductors (After [52.31] p. 118)

Compound	Symmetry	<i>a</i> (nm) <i>b</i> (nm)	<i>c</i> (nm)
La _{1.85} Sr _{0.15} CuO ₄	tetragonal	0.3779	1.3200
YBa ₂ Cu ₃ O _{6.9}	orthorhombic	0.3822 0.3891	1.1677
Bi _{2.2} Sr ₂ Ca _{0.8} Cu ₂ O _{8+δ} ^a	orthorhombic	0.5414 0.5418	3.089
(Bi,Pb) ₂ Sr _{1.72} Ca ₂ Cu ₃ O _{10+δ} ^a	orthorhombic	0.5392 0.5395	3.6985
Tl _{1.7} Ba ₂ Ca _{1.06} Cu _{2.32} O _{8+δ}	tetragonal	0.3857	2.939
Tl _{1.64} Ba ₂ Ca _{1.87} Cu _{3.11} O _{10+δ}	tetragonal	0.3822	3.626
Tl ₂ Ba ₂ Ca ₃ Cu ₄ O _{12-δ}	tetragonal	0.3850	4.1984
Tl _{1.1} Ba ₂ Ca _{0.9} Cu _{2.1} O _{7.1}	tetragonal	0.3851	1.2728
Tl _{1.1} Ba ₂ Ca _{1.8} Cu _{3.0} O _{9.7}	tetragonal	0.3843	1.5871
TlBa ₂ Ca ₃ Cu ₄ O ₁₁	tetragonal	0.3848	1.9001
HgBa ₂ CuO _{4+δ}	tetragonal	0.380	0.9509
HgBa ₂ CaCu ₂ O _{6+δ}	tetragonal	0.3859	1.2657
HgBa ₂ Ca ₂ Cu ₃ O _{8+δ}	tetragonal	0.3853	1.5818
HgBa ₂ Ca ₃ Cu ₄ O _{10+δ}	tetragonal	0.3854	1.9006

^a Using a pseudo-tetragonal unit cell a lattice parameter $a(\text{tetragonal}) \cong (a + b)/2\sqrt{2} \approx 0.38 \text{ nm}$ would result

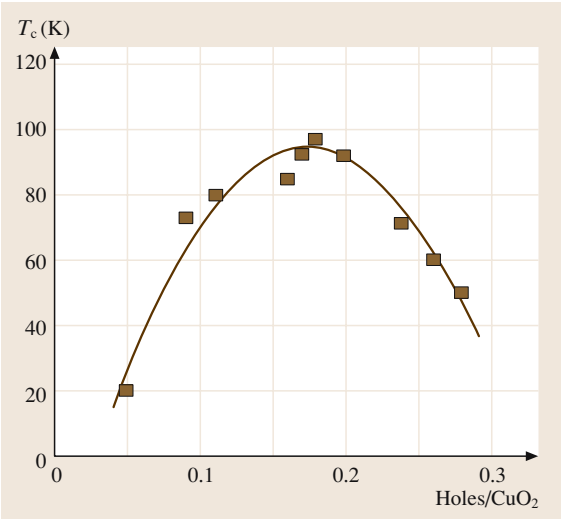


Fig. 52.20 Critical temperature of HgBa₂CuO_{4+δ} versus hole concentration. The maximum critical temperature is reached for a hole concentration of 0.18 holes/CuO₂ (After [52.31] p. 123)

In addition to the carrier concentration, the number of CuO₂ planes in the perovskite blocks is of importance for the critical temperature. The maximum critical temperatures of selected cuprate superconductors listed in Table 52.5 clearly indicate that the critical

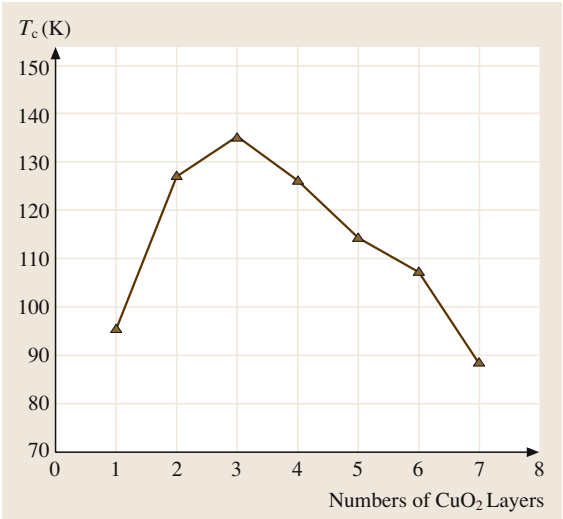


Fig. 52.21 Critical temperature of Hg-12(*n* − 1)*n* superconductors versus the number *n* of CuO₂ planes in the perovskite (copper oxide) blocks (After [52.48])

temperature increases with increasing number *n* of CuO₂ planes in the copper oxide blocks, at least up to the compound with *n* = 3. For compounds with four or more CuO₂ planes in the perovskite blocks,

the critical temperature decreases with increasing n . In Fig. 52.21, this behaviour is illustrated for the Hg-12($n-1$) n compounds. A possible explanation for the observed behaviour is the fact that in cuprates with $n \geq 3$ the hole concentrations in the central and outer CuO₂ layers may be considerably different. It may therefore be impossible to reach simultaneously the optimum carrier concentration in all the CuO₂ layers for compounds with $n \geq 4$.

Finally, it should be mentioned that the critical temperature of cuprate superconductors has been found to be

connected with the lattice parameters. Consequently, the critical temperature depends on applied pressure. Application of a hydrostatic pressure typically enhances the critical temperature of cuprate superconductors. In Hg-1223, superconductivity was observed at temperatures as high as 164 K at a pressure of 31 GPa [52.49]. A more complex situation results for the application of uniaxial pressure. In the case of pressure along the crystallographic a - or b -axis the critical temperature is enhanced, whereas for uniaxial pressure along the c -direction T_c is reduced [52.50, 51].

52.3 Physical Properties of Cuprate Superconductors

52.3.1 Anisotropic Superconductors

As a consequence of the layered crystal structures of the cuprates, the penetration depth depends on the direction of the screening currents. Assuming that the screening currents in the a - and b -direction are equal, two different penetration depths λ_{ab} and λ_c result. The indices give the direction of the screening currents. Thus, two different Ginzburg–Landau parameters $\kappa_{ab}(B||ab)$ and $\kappa_c(B||c)$ can be defined

$$\kappa_{ab} = \left(\frac{\lambda_{ab}\lambda_c}{\xi_{ab}\xi_c} \right)^{1/2}, \quad (52.20)$$

$$\kappa_c = \frac{\lambda_{ab}}{\xi_{ab}},$$

where ξ_{ab} and ξ_c are the coherence lengths in the ab -plane and along the c -direction respectively. The

Table 52.5 Critical temperatures of optimally doped cuprate superconductors

Compound	n	T_c (K)	Reference
Y-123	2	92	[52.20]
Bi-2212	2	96	[52.52]
Bi-2223	3	122	[52.53]
Tl-2201	1	92	[52.54]
Tl-2212	2	118	[52.54]
Tl-2223	3	128	[52.55]
Tl-2234	4	119	[52.53]
Tl-1212	2	90	[52.53]
Tl-1223	3	122	[52.53]
Tl-1234	4	122	[52.53]
Tl-1245	5	110	[52.53]
Hg-1223	3	135	[52.14]

dimensionless anisotropy parameter is given by [52.56]

$$\gamma_a = \left(\frac{m_c}{m_{ab}} \right)^{1/2} = \frac{\lambda_c}{\lambda_{ab}} = \frac{\xi_{ab}}{\xi_c} = \frac{B_{c2,ab}}{B_{c2,c}} = \frac{B_{c1,c}}{B_{c1,ab}} \quad (52.21)$$

where m_{ab} and m_c are the effective masses of the charge carriers for in-plane and out-of-plane currents, respectively. The expressions for the upper critical fields are [52.40, 56]

$$B_{c2,c} = \frac{\phi_0}{2\pi\xi_{ab}^2}$$

$$B_{c2,ab} = \frac{\phi_0}{2\pi\xi_c\xi_{ab}} \quad (52.22)$$

The upper and lower critical fields of various cuprate superconductors are listed in Table 52.6. The lower critical fields are typically between 10 and 500 mT for magnetic fields applied along the c -direction. The upper critical fields of cuprate superconductors can exceed 100 T even for the unfavourable field direction $B||c$. The data for Tl-2201, Hg-1201 and Hg, Cu-1201 given in Table 52.6 clearly indicate that the critical fields, like the critical temperatures, depend on the actual doping level.

For the characteristic length scales in cuprate superconductors the following relation has been found

$$\xi_c < \xi_{ab} \ll \lambda_{ab} < \lambda_c. \quad (52.23)$$

Table 52.7 lists for various cuprate superconductors the penetration depths and the coherence lengths. The in-plane coherence length ξ_{ab} is typically 1–4 nm. The considerably smaller values found for the out-of-plane coherence length ξ_c are comparable to interatomic distances. The small values of ξ_c are responsible for the insulating character of the charge-carrier reservoirs (Fig. 52.14) separating adjacent CuO₂ blocks. The

Table 52.6 Lower and upper critical fields of various cuprate superconductors

Compound	$B_{c1,c}$ (mT)	$B_{c2,c}$ (T) ^a	Comments	Reference
La, Sr-214		53		[52.57]
La-214		36/50	Ba/Sr doped	[52.58]
Y-123		122	$B_{c2,ab} = 674$ T	[52.59]
Y-123	500	29	$B_{c2,ab} = 140$ T	[52.60]
Y-123	150		$B_{c1,ab} = 24$ mT	[52.61]
Y-123	85		$B_{c1,ab} = 25$ mT	[52.62]
Bi-2201		20.2		[52.63]
Bi-2212		107		[52.64]
Bi-2212		31	$B_{c2,ab} = 476$ T	[52.65]
Bi-2212		71.5	$B_{c2,ab} = 542$ T	[52.66]
Bi, Pb-2212		83/95		[52.67]
Bi-2223		39	$B_{c2,ab} = 1210$ T	[52.68]
Bi, Pb-2223		184		[52.69]
Bi, Pb-2223	13.5			[52.70]
Tl-2201		0.85/1.9	$T_c = 9.7/15.7$ K	[52.71]
Tl-2212		99		[52.64]
Tl-2212	30		$B_{c1,ab} = 0.8$ mT	[52.61]
Tl-2223		75		[52.64]
Tl-1223	37	75		[52.72]
Hg-1201		100		[52.73]
Hg-1201	12.4/8.1		$T_c = 82/87$ K	[52.74]
Hg, Cu-1201		67/160	$T_c = 85/96$ K	[52.75]
Hg-1223		190		[52.34]

^a Including the upper critical field data of polycrystalline samples

cuprates are therefore nearly two-dimensional superconductors with intrinsic Josephson junctions along the *c*-direction. Moreover, the cuprates are extreme type II superconductors with κ_c values around 100.

52.3.2 Irreversibility Line

The magnetic phase diagram of cuprate superconductors is much more complicated than that of metallic superconductors. It is beyond the scope of this chapter to describe the details of the magnetic phase diagram of cuprate superconductors. A simplified magnetic phase diagram is shown in Fig. 52.22. An outstanding feature of high-temperature superconductors is the existence of an irreversibility line well below the upper critical field. Above this line the vortices are movable, and hence the critical current density is zero. The irreversibility line can be described by the expression

$$B_{irr} = B_0 \left(1 - \frac{T_{irr}}{T_c}\right)^\alpha$$

(52.24)

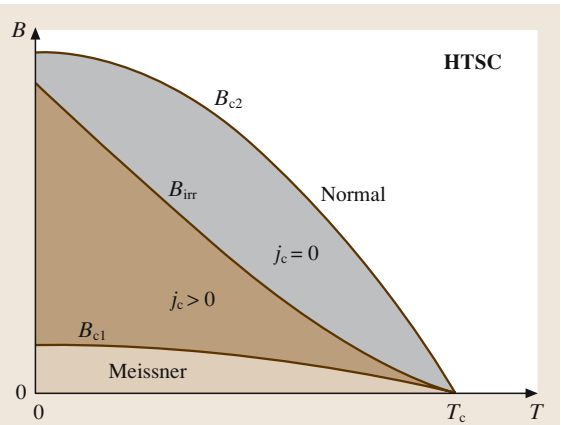


Fig. 52.22 Simplified magnetic phase diagram of cuprate superconductors. Above the irreversibility line the vortices are movable, and hence the critical current density is zero (After [52.31] p. 138)

where B_{irr} is the irreversibility field, T_{irr} the irreversibility temperature, B_0 a fitting parameter and the exponent

Table 52.7 Characteristic lengths for selected cuprate superconductors (data from [52.31, 34, 73, 75, 76])

Compound	λ_{ab} (nm)	ξ_{ab} (nm)	ξ_c (nm)
La, Sr-214	250–410	2.1	
Y-123	130–240	1.3–1.6	0.2–0.3
Bi-2201	438	4.0	
Bi-2212		2.7–3.8	0.16–0.18
Bi-2223		2.9	0.093
Bi, Pb-2212	178	2.0	
Bi, Pb-2223	88–245	1.35–2.9	
Tl-2201		5.2	0.3
Tl-2212		2.0–3.1	0.03–0.68
Tl-2223	117–163	1.1–1.36	
Tl-2234		4.5	1
Tl-1223	137	2.1	
Hg-1201	117–216*	1.7–3.4*	
Hg, Cu-1201	162–283*	1.43–2.2*	
Hg-1223	130–150	1.3	

* Different doping levels

α is characteristic of the superconductor in question. In the case of Y-123 the exponent α is close to 1.5 for magnetic fields perpendicular to the crystallographic c -direction, while values around 1.3 have been found for $B||c$ [52.61].

The existence of the irreversibility line is reflected in the broadening of the resistive transition in the presence of an applied magnetic field. Figure 52.23 shows the resistive transition of a Bi-2212 single crystal for magnetic

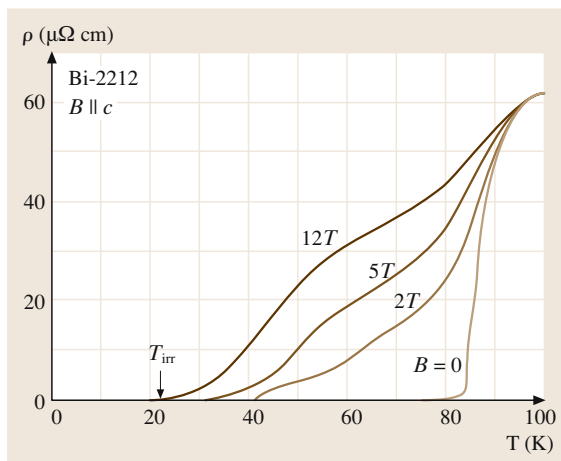


Fig. 52.23 Resistivity versus temperature curves of a Bi-2212 single crystal for magnetic fields parallel to the c -direction ([52.31] p. 139)

fields of 2, 5 and 12 T parallel to the c -direction. The data for zero applied field are also shown in Fig. 52.23. The end point of the resistive transition corresponds to the irreversibility temperature for the magnetic field applied. On the other hand, the onset of the resistivity drop may be used to determine the upper critical field.

The existence of an irreversibility line can be explained in two different ways. In the more conventional picture the occurrence of resistance above the irreversibility line is caused by thermally activated depinning. Within this picture it can be easily understood that the irreversibility line of cuprate superconductors is shifted to higher fields and temperatures by heavy-ion irradiation [52.77–79]. The irradiation generates amorphous tracks, which are efficient pinning centres. An alternative scenario is the melting of the flux-line lattice. Evidence for this possibility is the latent heat measured at the phase transition [52.80, 81].

The maximum operating field of cuprate superconductors is limited by the irreversibility line for the unfavourable field direction $B||c$. In any comparison of the behaviour of different cuprates the irreversibility field as a function of the reduced temperature T/T_c should be considered. The highest irreversibility fields result for Y-123, while the lowest values of B_{irr} have been found for Bi- and Tl-22($n-1$) n compounds. Intermediate irreversibility fields result for Tl- and Hg-12($n-1$) n compounds [52.31, 82]. The observed order in the irreversibility fields indicates that flux pinning is more efficient in cuprates with less anisotropic physical properties. The largest anisotropy of the physical properties has been observed in Bi- and Tl-22($n-1$) n compounds with the largest distance of adjacent CuO_2 blocks. The relatively small anisotropy of Y-123 seems to be the consequence of the short-circuiting of the insulating charge-carrier reservoirs by the CuO chains (Fig. 52.18).

52.3.3 Limitations of the Transport Critical Current

The electric field generated in a superconductor at the transition to the normal conducting state is typically well described by a power law $E = E_c(I/I_c)^n$, where I is the current, n is the index of resistive transition and E_c the electric field used to define the critical current I_c . Typically E_c values of 0.1 or 1 $\mu\text{V}/\text{cm}$ are used to determine the critical current.

It should be recognised that, unlike the critical temperature and the upper critical field, the critical current density is not an intrinsic material property. The

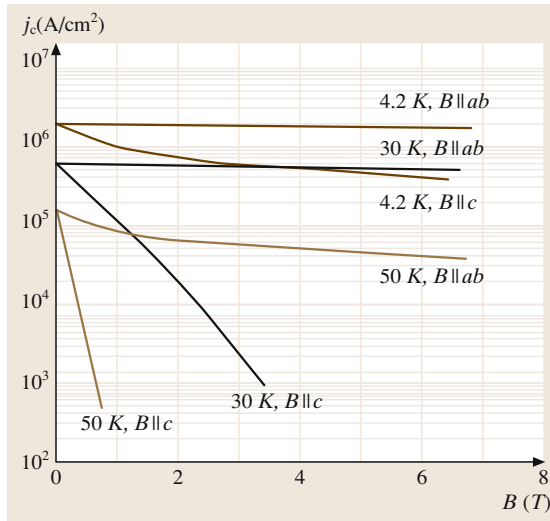


Fig. 52.24 Critical current densities in epitaxial Bi-2212 films as a function of applied magnetic field for B along the ab -planes and parallel to the crystallographic c -direction (After [52.83])

achievable j_c values are determined mainly by the microstructure of the superconductor in question. Grain boundaries can act as weak links, leading to small critical current densities in polycrystalline cuprate superconductors. In the absence of grain-boundary weak links, the transport j_c is mainly determined by flux pinning. Before these effects are discussed the anisotropy of j_c in individual grains and epitaxial thin films is considered.

Anisotropy of the Critical Current Density

In cuprate superconductors, conductive copper oxide blocks and insulating charge-carrier reservoirs alternate along the crystallographic c -direction. The currents can easily flow in the conductive CuO_2 planes, whereas currents along the c -direction have to cross the insulating or semiconducting charge carrier reservoirs (tunnelling). In spite of j_c values as high as 10^8 A/cm^2 in the CuO_2 planes of Hg-1212 films, the estimated j_c along the c -direction is as low as 5000 A/cm^2 [52.84].

A further anisotropy of the critical current density is related to the direction of the applied magnetic field. Figure 52.24 shows the critical current density of an epitaxial Bi-2212 film as a function of the applied magnetic field for selected temperatures. The critical current density declines rapidly with increasing applied field for $B \parallel c$, while the field dependence of j_c is much less pronounced for $B \parallel ab$. This behaviour reflects the considerably different upper critical fields for these two

field directions. The angle dependence of j_c in epitaxial Bi-2212 and Bi-2223 films can be described by the expression

$$j_c(B, \theta) = j_{c,c}(B \sin \theta), \quad (52.25)$$

where θ is the angle between the ab -planes and the field direction, and $j_{c,c}$ is the critical current density for $B \parallel c$. The above expression states that the reduction of the critical current by an applied magnetic field is determined exclusively by the field component parallel to the c -direction.

Grain-Boundary Weak Links

Soon after the discovery of superconductivity above 77 K in Y-123, it was noticed that the transport j_c in sintered samples is typically well below 1000 A/cm^2 . In addition, magnetic fields of a few mT are sufficient to reduce the j_c by an order of magnitude. On the other hand, critical current densities well above 10^6 A/cm^2 could be achieved in epitaxial Y-123 films. These experimental results suggest that the grain boundaries present in polycrystalline material can act as barriers for the transport current. Due to extremely short coherence lengths $\xi_c < 1 \text{ nm}$ the Cooper pair density can be considerably reduced within the width of a grain boundary. Two grains separated by a grain boundary can therefore be considered as a Josephson junction.

The high j_c values achieved in highly textured films suggest that low-angle grain boundaries are strongly coupled. To clarify this aspect, the transport properties of artificial grain boundaries have been studied [52.85,86]. In epitaxial growth the crystallography prescribed by the substrate can be transmitted to the Y-123 film. Using SrTiO_3 bicrystals, two Y-123 grains connected by an artificial grain boundary with a well defined misorientation angle θ can be prepared. The critical current density across the grain boundaries has been found to decrease rapidly with increasing misalignment angle. For [001]-tilt boundaries (rotation of the second grain around the common c -axis) in Y-123 films, the grain boundary j_c depends exponentially on the misalignment angle θ

$$j_c = j_{c0} \exp\left(-\frac{\theta}{\theta_0}\right), \quad (52.26)$$

where $\theta_0 \cong 6.3^\circ$ and $j_{c0} \cong 2 \times 10^7 \text{ A/cm}^2$ at 4.2 K [52.87]. The corresponding values for Bi-2223 [001]-tilt grain boundaries are $j_{c0} = 6 \times 10^6 \text{ A/cm}^2$ and $\theta_0 = 5^\circ$ at 26 K [52.88].

The weak-link behaviour is closely related to the strong reduction of the Cooper pair density in the grain

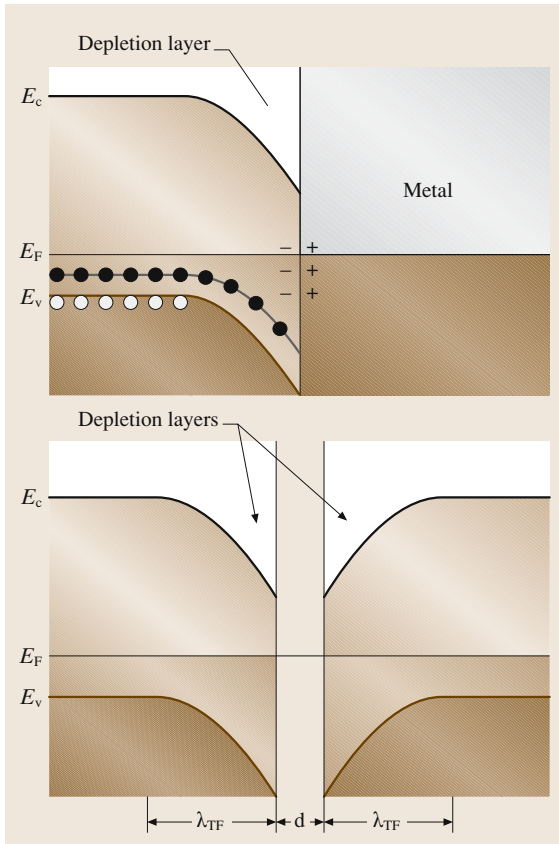


Fig. 52.25 Space charge effects and band bending at a semiconductor–metal interface (*top*) and at a grain boundary in a cuprate superconductor (*bottom*) (After [52.90])

boundaries caused by the extremely short coherence lengths and the low carrier concentrations in cuprate superconductors. The carrier concentration in optimally doped cuprate superconductors is around $5 \times 10^{21} \text{ cm}^{-3}$, whereas the carrier concentration in copper is as high as $85 \times 10^{21} \text{ cm}^{-3}$. Furthermore, the behaviour of grain boundaries in cuprate superconductors is similar to that of semiconductor–metal interfaces [52.89, 90]. In the case of a p-type semiconductor in contact with a metal characterised by a smaller work function, electrons flow from the metal into the semiconductor. As a consequence, both the valence and the conduction band are bent downwards (see Fig. 52.25 *top*). There results an excess of negatively charged acceptor ions and a depletion of holes in the valence band. Currents from the metal to the semiconductor would further enlarge the space-charge region and are therefore negligible (Schottky contact). Because of low carrier concentrations and

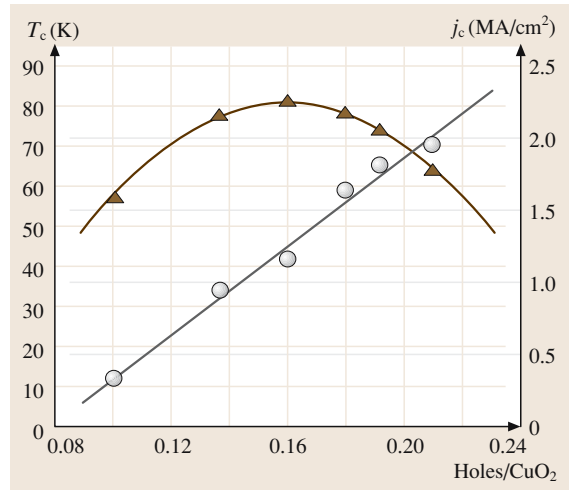


Fig. 52.26 Critical temperature (Δ) and current density (○) of a symmetric 24° [001]-tilt boundary as a function of the hole concentration. The charge-carrier concentration in the bicrystalline $\text{Y}_{0.7}\text{Ca}_{0.3}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ films has been modified by annealing at different oxygen pressures (After [52.87])

large dielectric constants ($\epsilon \approx 10$), band-bending effects can occur in cuprate superconductors. The grain boundaries of cuprate superconductors may be considered as double Schottky barriers (see Fig. 52.25 *bottom*). The resulting Thomas–Fermi screening length is in the range of 0.5–1 nm, which is comparable to the coherence length. The width of the insulating layer at the grain boundary is $t = d + 2l_d$, where d is the structural width of the boundary and l_d the thickness of the depletion layer. The structural width of typical grain boundaries is $\approx 0.4 \text{ nm}$. On the other hand, the thickness of the depletion layer is 0.16 nm for $\epsilon = 10$, an in-built voltage of 0.1 V and a carrier concentration of $4.5 \times 10^{21} \text{ cm}^{-3}$ [52.90]. Consequently, the total thickness of the insulating layer is considerably enhanced by the space-charge region.

In such a scenario it should be possible to modify the electronic properties of the interface by increasing the carrier concentration. In Y-123 this can be done by Ca doping. Figure 52.26 shows for symmetric 24° [001]-tilt boundaries the critical temperature and current density as a function of the hole concentration. As expected the j_c values at 4.2 K increase with increasing hole concentration even in the overdoped region, where T_c decreases.

Flux Pinning

Transport currents without losses can flow in a type II superconductor only when the flux lines are effectively

pinned. The screening currents circulating around each flux line are responsible for a repulsive vortex–vortex interaction, which prevents flux motion as soon as a sufficiently large fraction of the flux lines is pinned. The effectiveness of this collective pinning depends on the stiffness of the flux-line lattice.

The energy required to create a flux line can be estimated from the lower critical field B_{c1} [52.39]. At the lower critical field, flux lines start to enter the superconductor. The formation energy for each flux line is therefore

$$E_{fl} = \int_0^{B_{c1}} m dB \quad \text{with} \quad m = \frac{\phi_0}{\mu_0} l_{fl}, \quad (52.27)$$

where m is the magnetic moment of a flux line of length l_{fl} containing a single flux quantum. The required energy per unit length is

$$\frac{E_{fl}}{l_{fl}} = \frac{\phi_0}{\mu_0} B_{c1}. \quad (52.28)$$

Using $\kappa = \lambda_L/\xi$ the lower critical field can be written as [52.34]

$$B_{c1} = \frac{\phi_0 \ln \kappa}{4\pi\lambda_L^2}. \quad (52.29)$$

When the flux line is in a normal region, no energy is required to form the normal vortex core, and the resulting

pinning energy per unit length is

$$\frac{E_{fl}}{l_{fl}} = \frac{1}{4\pi\mu_0} \left(\frac{\phi_0}{\lambda_L} \right)^2 \ln \left(\frac{\lambda_L}{\xi} \right). \quad (52.30)$$

Generally regions with reduced Cooper pair densities due to defects can act as pinning centres.

The pinning properties of the cuprate superconductors are highly anisotropic. Relatively strong pinning has been found for magnetic fields transverse to the current direction applied in the ab plane. For the resulting Lorentz force along the c -direction, strong intrinsic pinning may be provided by the insulating charge carrier reservoirs of the layered structure along the c -direction. On the other hand, the pinning is weak for magnetic fields transverse to the current direction applied along the c -direction, leading to a Lorentz force parallel to the ab -planes. Flux lines along the c -direction experience strong pinning only within the CuO_2 blocks. As a consequence of the layered structure, the vortices may be cut into vortex segments or even pancake vortices which exist only within the conductive CuO_2 blocks. These pancake vortices could easily move in the direction of the Lorentz force.

The positions of the irreversibility lines in the different cuprate superconductors reflect the anisotropy of the pinning properties. Typically the pinning efficiency for magnetic fields along the c -direction decreases as one progresses from Y-123 via the Hg- and Tl- $12(n-1)n$ compounds to the highly anisotropic Bi- and Tl- $22(n-1)n$ cuprates.

52.4 Superconducting Films

Because of the weakly coupled large-angle grain boundaries present in polycrystalline bulk material, the transport critical current density is much smaller than the j_c of the individual grains. In contrast to polycrystalline bulk material, adjacent grains of epitaxial films are connected by low-angle grain boundaries. As a consequence the critical current densities measured in epitaxial films are close to the j_c values of individual grains. At low temperatures and zero applied field, the critical current densities in epitaxial Y-123, Bi-2212, Bi-2223 and Hg-1223 films exceed 10^7 A/cm^2 [52.83, 91–93]. Even at 77 K j_c -values well above 10^7 A/cm^2 have been measured in Y-123 films [52.94].

The physical properties of ultrathin films can be considerably different from those of bulk single crystals.

First, let us consider the effect of the film thickness on the critical temperature. For the growth of epitaxial ultrathin films a substrate with good lattice matching is required. $\text{YBa}_2\text{Cu}_3\text{O}_7$ films of thickness equal to the height of a single unit cell in the c -direction can be grown onto SrTiO_3 single crystals covered by a six-unit-cell-thick non-superconducting $\text{PrBa}_2\text{Cu}_3\text{O}_7$ (Pr-123) buffer layer [52.95]. The lattice mismatch of Y-123 and Pr-123 is less than 1.5%. The layer growth can be checked by the observation of the strong intensity oscillations of reflection high-energy electron diffraction. The period of these oscillations has been found to correspond to the height of a single unit cell. Y-123 films with a thickness of 1–10 unit cell heights were grown by coevaporation of the metals under oxygen atmosphere.

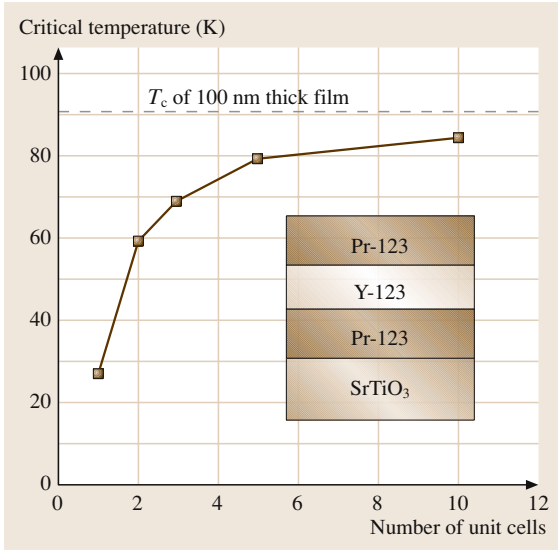


Fig. 52.27 Offset critical temperature of ultra-thin Y-123 films as a function of the layer thickness normalised to the c lattice parameter (After [52.95])

To provide hole doping the ultrathin Y-123 films are covered by a six-unit-cell-thick Pr-123 layer. The offset critical temperature of these Y-123 films is considerably lower than that of a 100-nm-thick film (Fig. 52.27). On the other hand, the onset critical temperature has been found to be much less affected. The reduction of the offset critical temperature has been attributed to the more pronounced critical fluctuations in the thinner films. Critical fluctuations are variations of the superconducting order parameter $\psi = n_C^{1/2} \exp(i\varphi)$ (n_C is the Cooper pair density) around the thermal equilibrium values.

The critical temperature of ultrathin Y-123 layers in Y-123/Pr-123 multilayers has been found to depend on the thicknesses of the insulating Pr-123 layers [52.96]. The thickness of 1.2 nm of the embedded Y-123 layers corresponds to the height of a single unit cell in the c -direction. The critical temperature of the Y-123/Pr-123 multilayer decreases from ≈ 50 K (10% of normal-state resistance) to ≈ 13 K for Pr-123 interlayer thicknesses of 1.2 and 7.2 nm, respectively. Because of the Josephson coupling of adjacent Y-123 layers in the multilayer the critical fluctuations are expected to be less pronounced than in a single Y-123 film of the same thickness. Thus, the T_c of the Y-123 layers in the Y-123/Pr-123 multilayer with a Pr-123 thickness of 1.2 nm is higher than that of a single Y-123 layer of the same thickness (Fig. 52.27).

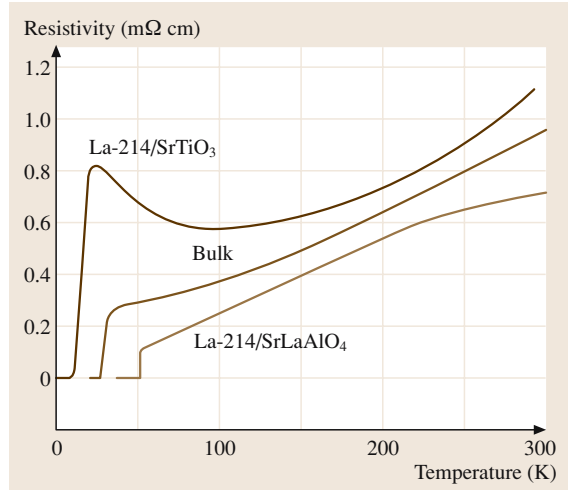


Fig. 52.28 Resistivity versus temperature for La-214 films on SrTiO₃ and SrLaAlO₄ substrates. The different epitaxial strains lead to considerably different critical temperatures of 10 and 49 K. For comparison the resistivity of La-214 bulk material ($T_c = 25$ K) is also shown (After [52.51])

Finally, let us consider the effects of epitaxial strain on the critical temperature of (La, Sr)₂CuO₄ films [52.51, 97, 98]. In epitaxial growth the crystallography prescribed by the substrate is transmitted to the film. For a sufficiently small film thickness the lattice constant of the film is equal to that of the substrate. As a consequence of the lattice mismatch a considerable epitaxial strain can be induced in the film. Under epitaxial growth conditions a compressive strain in the CuO₂ planes of (La, Sr)₂CuO₄ is accompanied by a tensile strain in the c -direction. Typically, in-plane compressive and out-of-plane tensile strains enhance the T_c of cuprate superconductors. Thus, the critical temperature of cuprate superconductors should be much more sensitive to epitaxial strain effects than to hydrostatic pressure, where the opposite in-plane and out-of-plane effects partly cancel each other. Because of film deposition at temperatures above 750 °C, differences in the thermal expansion coefficients provide an additional contribution to the strain in the films. In thick films the epitaxial strain is released by the appearance of misfit dislocations. The expected change in the critical temperature of the films is given by

$$T_c = T_c(0) + 2 \frac{dT_c}{d\varepsilon_{ab}} \varepsilon_{ab} + \frac{dT_c}{d\varepsilon_c} \varepsilon_c, \quad (52.31)$$

where $\varepsilon = (d_{\text{bulk}} - d_{\text{strained}})/d_{\text{bulk}}$, with d a lattice parameter.

To study the epitaxial strain effects, Locquet et al. deposited $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$ films ($a = 0.3784 \text{ nm}$, $\alpha = 8.5 \times 10^{-6} \text{ K}^{-1}$) of 10–15 nm thickness on SrTiO_3 ($a = 0.3905 \text{ nm}$, $\alpha = 9 \times 10^{-6} \text{ K}^{-1}$) and SrLaAlO_4 ($a = 0.3754 \text{ nm}$, $\alpha = 10.5 \times 10^{-6} \text{ K}^{-1}$) substrates. Using X-ray diffraction, strain values of $\varepsilon_{ab} = 0.63\%$ and $\varepsilon_c = -0.76\%$ have been determined for the La-214 film on SrTiO_3 . The corresponding strain values for the La-214 film deposited onto SrLaAlO_4 are $\varepsilon_{ab} = -0.54\%$ and $\varepsilon_c = 0.35\%$. The results of Locquet et al. are

shown in Fig. 52.28. The critical temperatures of the La-214 films of 10 (SrTiO_3) and 49 K (SrLaAlO_4) are, as expected, considerably different from the bulk value of 25 K. In addition, the temperature dependencies of the normal state resistivities of the two films are remarkably different. The film experiencing a tensile in-plane strain shows an upturn of the normal state resistance at temperatures below 100 K, which is an effect also found in underdoped cuprate superconductors.

52.5 The Special Case of MgB_2

In spite of the fact that MgB_2 has been known since the early 1950s, it was unknown until 2001 that it is a superconductor with a remarkably high transition temperature. Figure 52.29 shows the hexagonal crystal structure of MgB_2 , consisting of alternating layers of boron and magnesium. As a consequence of the layered structure, anisotropic physical properties result. The arrangement of the boron atoms is the same as that of the carbon atoms in graphite. Each of the boron atoms is surrounded by three other boron atoms at the same distance.

To study the role of phonons in the superconductivity of MgB_2 , the critical temperatures of isotopically pure Mg^{10}B_2 and Mg^{11}B_2 specimens have been studied [52.99]. The transition temperatures of Mg^{11}B_2 and Mg^{10}B_2 defined by a 2% onset criterion are 39.2 and

40.2 K, respectively. Defining a partial boron isotope-effect exponent $\beta_B = \Delta \ln T_c / \Delta \ln M_B$, where M_B is the atomic mass of boron, a value of $\beta_B = 0.26 \pm 0.03$ results from the experimental data. On the other hand, the isotope effect for $^{26}\text{MgB}_2$ and $^{24}\text{MgB}_2$ is much less pronounced [52.100]. The shift of 0.1 K in T_c corresponds to a partial Mg isotope-effect exponent β_{Mg} of 0.02. The strong boron isotope effect suggests that the pairing of the electrons is caused by the boron phonons. Thus, MgB_2 seems to be a conventional metallic superconductor with an extremely high transition temperature. The high transition temperature seems to be caused by the small atomic mass of boron, leading to high phonon frequencies.

The currently known physical properties of MgB_2 are summarised in Table 52.8. As in cuprate superconductors the operation range of MgB_2 is limited by the irreversibility field. In polycrystalline bulk samples the irreversibility field typically reaches 50% of B_{c2} . In nanocrystalline MgB_2 bulk material the irreversibility field is shifted to 80% of B_{c2} . The enhancement of the irreversibility field in the nanocrystalline material has been attributed to the improved pinning provided by the large number of grain boundaries [52.101]. Maximum irreversibility fields of more than 14 T at 4.2 K have been reported for MgB_2 bulk material annealed in Mg vapour [52.102]. MgB_2 is an extreme type II superconductor with a Ginzburg–Landau parameter κ of 38. The coherence length of $\xi_c \approx 2.5 \text{ nm}$ is considerably larger than the values of $\approx 0.2 \text{ nm}$ measured for cuprate superconductors (Table 52.7). The anisotropy of the upper critical fields reported for c -axis-oriented films [52.103] and single crystals [52.104] is typically close to $B_{c2,ab}/B_{c2,c} \approx 3$.

Immediately after the discovery of superconductivity in MgB_2 the current transport across grain boundaries

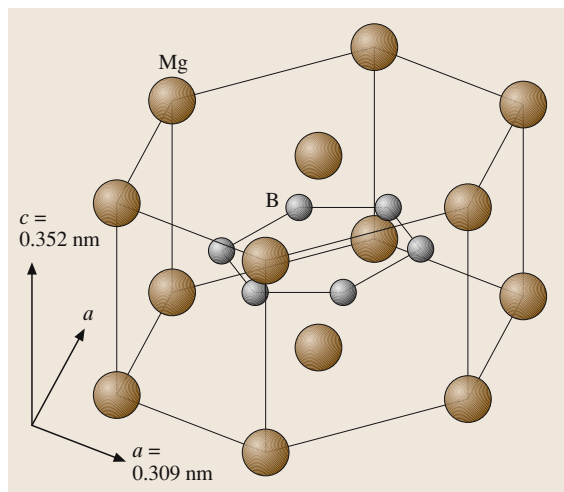


Fig. 52.29 Hexagonal crystal structure of MgB_2 consisting of alternating layers of magnesium and boron (After [52.15])

Table 52.8 Physical properties of MgB₂

Parameter	Value	Comments	Reference
T_c (K)	40.2	Mg ¹⁰ B ₂	[52.99]
Lattice parameters		Crystal structure	[52.15]
a (nm)	0.3086	Hexagonal	
c (nm)	0.3524	Space group P6/mmm	
ρ (mWcm)	1–18	at 40 K	[52.102]
$\rho(300\text{ K})/\rho(40\text{ K})$	15–3		[52.102]
B_{c1} (mT)	18–26.5	$T = 0$	[52.107, 108]
$B_{c2,ab}$ (T)	19.6/26.4	Films T_c : 37.5/31.4 K, $T = 0$	[52.109]
$B_{c2,c}$ (T)	13.8/14.6	Films T_c : 37.5/31.4 K, $T = 0$	[52.109]
B_{irr} (T)	$> 14/\approx 6$	$T = 4.2/20\text{ K}$	[52.102]
$\chi_a b$ (nm)	6.8–8.1	$T = 0$	[52.108, 110]
χ_c (nm)	2.3–2.7	$T = 0$	[52.108, 110]
λ (nm)	110–180	$T = 0$	[52.107, 111]
$\kappa = \lambda/\chi$	38	polycrystalline, $T = 0$	[52.107]
2Δ (meV)	17–19*	$T = 4.2\text{ K}$	[52.112]

* MgB₂ seems to be a double gap superconductor (see for example [52.113])

was studied [52.105, 106]. The results of these investigations indicate that the grain boundaries in MgB₂ do not act as weak links. This behaviour reflects the fact that the coherence length of MgB₂ is considerably larger than that of cuprate superconductors.

To give an impression of the current-carrying capacity of MgB₂ the critical current density of c -axis-oriented films as a function of temperature is shown in Fig. 52.30 for selected applied magnetic fields ($B||c$). The critical current density at 15 K and zero applied field is as high as 16 MA/cm² [52.114, 115], which is comparable to the j_c values of cuprate superconductors.

Finally, let us briefly consider the synthesis of MgB₂. Polycrystalline MgB₂ can be synthesised by the reaction of boron powder with magnesium vapour. Stoichiometric quantities of boron and magnesium are sealed into a tantalum tube under an argon atmosphere. In addition, the tantalum tube containing the B/Mg mixture is sealed into a quartz ampoule. To form the desired MgB₂ intermetallic compound the mixture is heat treated at 950 °C for 2 h [52.116]. In a similar way boron precursor films can be converted to MgB₂ in the presence of Mg vapour. The annealing temperature for the formation of MgB₂ films is typically in the 850 °C range. Very recently c -axis-oriented epitaxial MgB₂ films have been grown onto Al₂O₃ substrates [52.103]. For applications in electronics it is

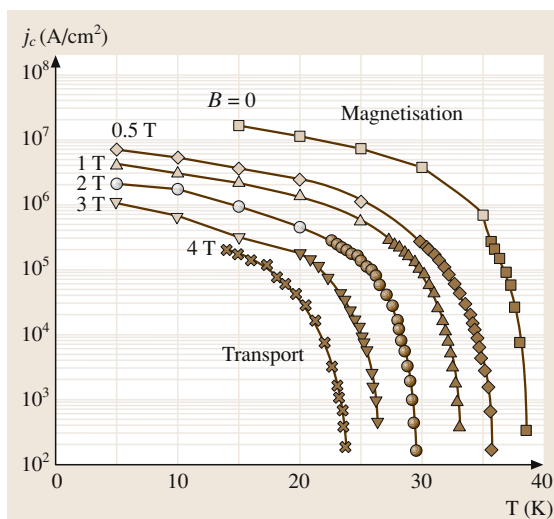


Fig. 52.30 Critical current density of a c -axis-oriented MgB₂ film as a function of temperature for magnetic fields up to 4 T (After [52.114])

desirable to grow the MgB₂ films in situ in a single step on a suitable substrate. The substrate and electronic materials ZrO₂, MgO, TiN, TaN, AlN and SiC show no reaction with MgB₂ up to an annealing temperature of 800 °C [52.117].

52.6 Summary

The cuprate high-temperature superconductors are characterised by layered crystal structures. Conductive CuO₂ layers in the crystallographic *ab*-planes are a common feature of all cuprate superconductors. Along the *c*-direction, conducting copper oxide blocks and insulating charge-carrier reservoirs alternate. As a consequence the physical properties of cuprate superconductors are strongly anisotropic.

The critical temperature as well as the other physical properties depend considerably on the charge-carrier concentration. The maximum critical temperature is typically reached for a carrier concentration of 0.18 holes per CuO₂. Furthermore, the highest T_c of optimally doped cuprate superconductors is typically reached in compounds with three or four CuO₂ layers in the CuO₂ blocks. The highest known critical temperature of 135 K at ambient pressure has been achieved in Hg-1223.

The cuprate superconductors are close to a metal–insulator transition. The d-wave symmetry of the energy gap and the existence of a pseudogap above T_c , especially in underdoped cuprate superconductors, suggest that the pairing mechanism of the electrons is different from that in metallic superconductors. Furthermore,

only a weak isotope effect has been observed in the cuprates, which again suggests that the electron pairing is not phonon-mediated.

Extremely short coherence lengths and low carrier concentrations seem to be responsible for the weak-link behaviour of large-angle grain boundaries in the cuprate superconductors. To achieve transport critical current densities comparable to the intra-grain j_c a biaxial texture is required.

The recently discovered metallic superconductor MgB₂ may be interesting for applications in spite of the moderate critical temperature of 39 K because of a simpler chemistry and the fact that grain boundaries do not act as weak links. The observation of a pronounced boron isotope effect suggests that MgB₂ is a conventional BCS superconductor. The anisotropy of the physical properties of MgB₂ is less pronounced than that of the cuprate superconductors.

The maximum operation fields of both the cuprate superconductors and MgB₂ are limited by the irreversibility line, which is well below the upper critical field. The existence of an irreversibility line may be caused in both types of materials by thermally activated depinning or flux-line lattice melting.

References

- 52.1 R. de Bruyn Ouboter: IEEE Trans. Magn. **23**, 355 (1987)
- 52.2 W. Meissner, R. Ochsenfeld: Naturwissenschaften **21**, 787 (1933)
- 52.3 F. London, H. London: Physica **2**, 341 (1935)
- 52.4 J. Bardeen, L. N. Cooper, J. R. Schrieffer: Phys. Rev. **106**, 162 (1957)
- 52.5 J. Bardeen, L. N. Cooper, J. R. Schrieffer: Phys. Rev. **108**, 1175 (1957)
- 52.6 I. Giaever: Phys. Rev. Lett. **5**, 147 (1960)
- 52.7 I. Giaever: Phys. Rev. Lett. **5**, 464 (1960)
- 52.8 R. Doll, M. Näbauer: Phys. Rev. Lett. **7**, 51 (1961)
- 52.9 B. S. Deaver Jr., W. M. Fairbank: Phys. Rev. Lett. **7**, 43 (1961)
- 52.10 B. D. Josephson: Phys. Lett. **1**, 251 (1962)
- 52.11 R. C. Jaklevic, J. Lambe, J. E. Mercereau, A. H. Silver: Phys. Rev. A **140**, 1628 (1965)
- 52.12 J. G. Bednorz, K. A. Müller: Z. Phys. B **64**, 189 (1986)
- 52.13 A. Schilling, M. Cantoni, J. D. Guo, H. R. Ott: Nature **363**, 56 (1993)
- 52.14 J. J. Capponi, J. L. Tholence, C. Chaillout, M. Marezio, P. Bordet, J. Chenavas, S. M. Loureiro, E. V. Antipov, E. Kopnine, M. F. Gorius, M. Nunez-Regueiro, B. Souletie, P. Radaelli, F. Gerhards: Physica C **235–240**, 146 (1994)
- 52.15 J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zen-itani, J. Akimitsu: Nature **410**, 63 (2001)
- 52.16 L. R. Testardi, J. H. Wernick, W. A. Royer: Solid State Commun. **15**, 1 (1974)
- 52.17 J. R. Gavaler, M. A. Janocko, C. K. Jones: J. Appl. Phys. **45**, 3009 (1974)
- 52.18 J. F. Schooley, W. R. Hosler, M. L. Cohen: Phys. Rev. Lett. **12**, 474 (1964)
- 52.19 A. W. Sleight, J. L. Gillson, P. E. Bierstedt: Solid State Commun. **17**, 27 (1975)
- 52.20 M. K. Wu, J. R. Ashburn, C. J. Torng, P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, Y. Q. Wang, C. W. Chu: Phys. Rev. Lett. **58**, 908 (1987)
- 52.21 T. R. Strobridge: IEEE Trans. Nucl. Sci. **16**, 1104 (1969)
- 52.22 D. Jérôme, A. Mazaud, M. Ribault, K. Bechgaard: J. Physique Lett. **41**, L95 (1980)
- 52.23 K. Oshima, H. Urayama, H. Yamochi, G. Saito: Physica C **153–155**, 1148 (1988)
- 52.24 J. M. Williams, A. J. Schultz, U. Geiser, K. D. Carlson, A. M. Kini, H. H. Wang, W.-K. Kwok, M.-H. Whangbo, J. E. Schirber: Science **252**, 1501 (1991)

- 52.25 A. F. Hebard, M. J. Rosseinsky, R. C. Haddon, D. W. Murphy, S. H. Glarum, T. T. M. Palstra, A. P. Ramirez, A. R. Kortan: *Nature* **350**, 600 (1991)
- 52.26 K. Lüders: *Phys. Bl.* **50**, 166 (1994)
- 52.27 T. T. M. Palstra, O. Zhou, Y. Iwasa, P. E. Sulewski, R. M. Fleming, B. R. Zegarski: *Solid State Commun.* **93**, 327 (1995)
- 52.28 W.-J. Yeh, L. Chen, F. Xu, B. Bi, P. Yang: *Phys. Rev. B* **36**, 2414 (1987)
- 52.29 F. J. Kedves, S. Mészáros, K. Vad, G. Halász, B. Keszei, L. Mihály: *Solid State Commun.* **63**, 991 (1987)
- 52.30 D. J. Quinn, W. B. Ittner: *J. Appl. Phys.* **33**, 748 (1962)
- 52.31 R. Wesche: *High-temperature Superconductors: Materials, Properties, and Applications* (Kluwer Academic, Boston 1998)
- 52.32 H. Fröhlich: *Phys. Rev.* **79**, 845 (1950)
- 52.33 J. Bardeen: *Phys. Rev.* **80**, 567 (1950)
- 52.34 C. P. Poole, H. A. Farach, R. J. Creswick: *Superconductivity* (Academic, San Diego 1995)
- 52.35 C. A. Reynolds, B. Serin, W. H. Wright, L. B. Nesbitt: *Phys. Rev.* **78**, 487 (1950)
- 52.36 E. Maxwell: *Phys. Rev.* **78**, 477 (1950)
- 52.37 B. Serin, C. A. Reynolds, C. Lohman: *Phys. Rev.* **86**, 162 (1952)
- 52.38 J. W. Garland: *Phys. Rev. Lett.* **11**, 114 (1963)
- 52.39 W. Buckel: *Supraleitung* (VCH, Weinheim 1993)
- 52.40 T. P. Orlando, K. A. Delin: *Foundations of applied superconductivity* (Addison-Wesley, Reading 1991)
- 52.41 G. Sparr, J. D. Thompson, R. L. Whetten, S.-M. Huang, R. B. Kaner, F. Diederich, G. Grüner, K. Holczer: *Phys. Rev. Lett.* **68**, 1228 (1992)
- 52.42 Ch. Renner, I. Maggio-Aprile, Ø. Fischer: *Phys. Bl.* **54**, 427 (1998)
- 52.43 Ch. Renner, B. Revaz, J.-Y. Genoud, K. Kadowaki, Ø. Fischer: *Phys. Rev. Lett.* **80**, 149 (1998)
- 52.44 J. Orenstein, A. J. Millis: *Science* **288**, 468 (2000)
- 52.45 J. M. Tranquada, B. J. Sternlieb, J. D. Axe, Y. Nakamura, S. Uchida: *Nature* **375**, 561 (1995)
- 52.46 R. Beyers, G. Lim, E. M. Engler, R. J. Savoy, T. M. Shaw, T. R. Dinger, W. J. Gallagher, R. L. Sandstrom: *Appl. Phys. Lett.* **50**, 1918 (1987)
- 52.47 R. J. Cava, B. Batlogg, R. B. van Dover, D. W. Murphy, S. Sunshine, T. Siegrist, J. P. Remeika, E. A. Rietman, S. Zahurak, G. P. Espinosa: *Phys. Rev. Lett.* **58**, 1676 (1987)
- 52.48 J. J. Capponi, E. M. Kopnin, S. M. Loureiro, E. V. Antipov, E. Gautier, C. Chaillout, B. Souletie, M. Brunner, J. L. Tholence, M. Marezio: *Physica C* **256**, 1 (1996)
- 52.49 L. Gao, Y. Y. Xue, F. Chen, Q. Xiong, R. L. Meng, D. Ramirez, C. W. Chu, J. H. Eggert, H. K. Mao: *Phys. Rev. B* **50**, 4260 (1994)
- 52.50 N. Watanabe, K. Fukamachi, Y. Ueda, K. Tsushima, A. M. Balbashov, T. Nakanishi, N. Môri: *Physica C* **235-240**, 1309 (1994)
- 52.51 J.-P. Locquet, J. Perret, J. Fompeyrine, E. Mächler, J. W. Seo, G. van Tendeloo: *Nature* **394**, 453 (1998)
- 52.52 Th. Schweizer, R. Müller, L. J. Gauckler: *Physica C* **225**, 143 (1994)
- 52.53 R. Hott, H. Rietschel, M. Sander: *Phys. Bl.* **48**, 355 (1992)
- 52.54 A. Maignan, C. Martin, M. Huve, J. Provost, M. Hervieu, C. Michel, B. Raveau: *Physica C* **170**, 350 (1990)
- 52.55 A. Maignan, C. Martin, V. Hardy, Ch. Simon, M. Hervieu, B. Raveau: *Physica C* **219**, 407 (1994)
- 52.56 M. Tinkham: *Physica C* **235-240**, 3 (1994)
- 52.57 A. J. Panson, G. R. Wagner, A. I. Braginski, J. R. Gavaler, M. A. Janocko, H. C. Pohl, J. Talvacchio: *Appl. Phys. Lett.* **50**, 1104 (1987)
- 52.58 D. W. Capone, D. G. Hinks, J. D. Jorgensen, K. Zhang: *Appl. Phys. Lett.* **50**, 543 (1987)
- 52.59 U. Welp, W. K. Kwok, G. W. Crabtree, K. G. Vandervoort, J. Z. Liu: *Phys. Rev. Lett.* **62**, 1908 (1989)
- 52.60 T. K. Worthington, W. J. Gallagher, T. R. Dinger: *Phys. Rev. Lett.* **59**, 1160 (1987)
- 52.61 Ch. Heinzel, Ch. Neumann, P. Ziemann: *Europhys. Lett.* **13**, 531 (1990)
- 52.62 D.-H. Wu, S. Sridhar: *Phys. Rev. Lett.* **65**, 2074 (1990)
- 52.63 M. Akamatsu, L. X. Chen, H. Ikeda, R. Yoshizaki: *Physica C* **235-240**, 1619 (1994)
- 52.64 D. R. Harshman, A. P. Mills: *Phys. Rev. B* **45**, 10684 (1992)
- 52.65 J. H. Kang, R. T. Kampwirth, K. E. Gray: *Appl. Phys. Lett.* **52**, 2080 (1988)
- 52.66 J. N. Li, K. Kadowaki, M. J. V. Menken, Y. K. Huang, K. Bakker, A. A. Menovsky, J. J. M. Franse: *Appl. Phys. A* **47**, 209 (1988)
- 52.67 L. Zhang, J. Z. Liu, R. N. Shelton: *Phys. Rev. B* **45**, 4978 (1992)
- 52.68 I. Matsubara, H. Tanigawa, T. Ogura, H. Yamashita, M. Kinoshita, T. Kawai: *Phys. Rev. B* **45**, 7414 (1992)
- 52.69 F. Shi, T. S. Rong, S. Z. Zhou, X. F. Wu, J. Du, Z. H. Shi, C. G. Cui, R. Y. Jin, J. L. Zhang, Q. Z. Ran, N. C. Shi: *Phys. Rev. B* **41**, 6541 (1990)
- 52.70 I. Matsubara, R. Funahashi, K. Ueno, H. Yamashita, T. Kawai: *Physica C* **256**, 33 (1996)
- 52.71 M. Hasegawa, K. Izawa, A. Shibata, Y. Matsuda: *Physica C* **377**, 459 (2002)
- 52.72 G. Triscone, A. Junod, R. E. Gladyshevskii: *Physica C* **264**, 233 (1996)
- 52.73 J. Hofer, J. Karpinski, M. Willemin, G. I. Meijer, E. M. Kopnin, R. Molinski, H. Schwer, C. Rossel, H. Keller: *Physica C* **297**, 103 (1998)
- 52.74 R. Usami, A. Fukuoka, H. Kubota, H. Yamauchi: *Physica C* **243**, 19 (1995)
- 52.75 G. Villard, A. Daignere, D. Pelloquin, A. Maignan: *Physica C* **314**, 196 (1999)
- 52.76 V. Vulcanescu, L. Fruchter, A. Bertinotti, D. Colson, G. Le Bras, J.-F. Marucco: *Physica C* **259**, 131 (1996)
- 52.77 H. Kumakura, K. Togano, N. Tomita, E. Yanagisawa, S. Okayasu, Y. Kazumatu: *Physica C* **251**, 231 (1995)
- 52.78 V. Hardy, J. Provost, D. Groult, M. Hervieu, B. Raveau, S. Durčok, E. Pollert, J. C. Frison,

- J. P. Chaminade, M. Pouchard: *Physica C* **191**, 85 (1992)
- 52.79 L. Civale, A. D. Marwick, R. Wheeler, M. A. Kirk, W. L. Carter, G. N. Riley, A. P. Malozemoff: *Physica C* **208**, 137 (1993)
- 52.80 A. Schilling, R. A. Fisher, N. E. Phillips, U. Welp, D. Dasgupta, W. K. Kwok, G. W. Crabtree: *Nature* **382**, 791 (1996)
- 52.81 A. Junod, M. Roulin, J.-Y. Genoud, B. Revaz, A. Erb, E. Walker: *Physica C* **275**, 245 (1997)
- 52.82 M. Rupp, A. Gupta, C. C. Tsuei: *Appl. Phys. Lett.* **67**, 291 (1995)
- 52.83 P. Schmitt, L. Schultz, G. Saemann-Ischenko: *Physica C* **168**, 475 (1990)
- 52.84 L. Krusin-Elbaum, C. C. Tsuei, A. Gupta: *Nature* **373**, 679 (1995)
- 52.85 D. Dimos, P. Chaudhari, J. Mannhart, F. K. LeGoues: *Phys. Rev. Lett.* **61**, 219 (1988)
- 52.86 D. Dimos, P. Chaudhari, J. Mannhart: *Phys. Rev. B* **41**, 4038 (1990)
- 52.87 A. Schmehl, B. Goetz, R. R. Schulz, C. W. Schneider, H. Bielefeldt, H. Hilgenkamp, J. Mannhart: *Europhys. Lett.* **47**, 110 (1999)
- 52.88 A. Attenberger, J. Hänisch, B. Holzapfel, L. Schultz: *Physica C* **372–376**, 649 (2002)
- 52.89 H. Hilgenkamp, J. Mannhart: *Appl. Phys. Lett.* **73**, 265 (1998)
- 52.90 J. Mannhart, H. Hilgenkamp: *Physica C* **317–318**, 383 (1999)
- 52.91 R. A. Rao, Q. Gan, C. B. Eom, Y. Suzuki, A. A. McDaniel, J. W. P. Hsu: *Appl. Phys. Lett.* **69**, 3911 (1996)
- 52.92 H. Yamasaki, K. Endo, S. Kosaka, M. Umeda, S. Misawa, S. Yoshida, K. Kajimura: *IEEE Trans. Appl. Supercond.* **3**, 1536 (1993)
- 52.93 S. H. Yun, J. Z. Wu: *Appl. Phys. Lett.* **68**, 862 (1996)
- 52.94 D. Kumar, M. Sharon, R. Pinto, P. R. Apte, S. P. Pai, S. C. Purandare, L. C. Gupta, R. Vijayaraghavan: *Appl. Phys. Lett.* **62**, 3522 (1993)
- 52.95 T. Terashima, K. Shimura, Y. Bando, Y. Matsuda, A. Fujiyama, S. Komiyama: *Phys. Rev. Lett.* **67**, 1362 (1991)
- 52.96 J.-M. Triscone, Ø. Fischer, O. Brunner, L. Antognazza, A. D. Kent, M. G. Karkut: *Phys. Rev. Lett.* **64**, 804 (1990)
- 52.97 W. Si, H.-C. Li, X. X. Xi: *Appl. Phys. Lett.* **74**, 2839 (1999)
- 52.98 T. Sekitani, H. Sato, M. Naito, N. Miura: *Physica C* **378–381**, 195 (2002)
- 52.99 S. L. Bud'ko, G. Lapertot, C. Petrovic, C. E. Cunningham, N. Anderson, P. C. Canfield: *Phys. Rev. Lett.* **86**, 1877 (2001)
- 52.100 D. G. Hinks, H. Claus, J. D. Jorgensen: *Nature* **411**, 457 (2001)
- 52.101 A. Gümbel, J. Eckert, G. Fuchs, K. Nenkov, K.-H. Müller, L. Schultz: *Appl. Phys. Lett.* **80**, 2725 (2002)
- 52.102 V. Braccini, L. D. Cooley, S. Patnaik, D. C. Larbalestier, P. Manfrinetti, A. Palenzona, A. S. Siri: *Appl. Phys. Lett.* **81**, 4577 (2002)
- 52.103 S. D. Bu, D. M. Kim, J. H. Choi, J. Giencke, E. E. Hellstrom, D. C. Larbalestier, S. Patnaik, L. Cooley, C. B. Eom, J. Lettieri, D. G. Schlom, W. Tian, X. Q. Pan: *Appl. Phys. Lett.* **81**, 1851 (2002)
- 52.104 T. Masui, S. Lee, A. Yamamoto, S. Tajima: *Physica C* **378–381**, 216 (2002)
- 52.105 D. C. Larbalestier, L. D. Cooley, M. O. Rikel, A. A. Polyanskii, J. Jiang, S. Patnaik, X. Y. Cai, D. M. Feldmann, A. Gurevich, A. A. Squitieri, M. T. Naus, C. B. Eom, E. E. Hellstrom, R. J. Cava, K. A. Regan, N. Rogado, M. A. Hayward, T. He, J. S. Slusky, P. Khalifah, K. Inumaru, M. Haas: *Nature* **410**, 186 (2001)
- 52.106 K. H. P. Kim, W. N. Kang, M.-S. Kim, C. U. Jung, H.-J. Kim, E.-M. Choi, M.-S. Park, S.-I. Lee: *Physica C* **370**, 13 (2002)
- 52.107 Y. Wang, T. Plackowski, A. Junod: *Physica C* **355**, 179 (2001)
- 52.108 Z. X. Shi, A. K. Pradhan, M. Tokunaga, K. Yamazaki, T. Tamegai, Y. Takano, K. Togano, H. Kito, H. Ihara: *Physica C* **378–381**, 550 (2002)
- 52.109 C. Ferdeghini, V. Ferrando, G. Grassano, W. Ramadan, V. Braccini, M. Putti, P. Manfrinetti, A. Palenzona: *Physica C* **372–376**, 1270 (2002)
- 52.110 Y. Eltsev, S. Lee, K. Nakao, N. Chikumoto, S. Tajima, N. Koshizuka, M. Murakami: *Physica C* **378–381**, 61 (2002)
- 52.111 F. Simon, A. Jánossy, T. Fehér, F. Murányi, S. Garaj, L. Forró, C. Petrovic, S. L. Bud'ko, G. Lapertot, V. G. Kogan, P. C. Canfield: *Phys. Rev. Lett.* **87**, 047002 (2001)
- 52.112 T. Takasaki, T. Ekino, T. Muranaka, H. Fujii, J. Akimitsu: *Physica C* **378–381**, 229 (2002)
- 52.113 F. Bouquet, R. A. Fisher, N. E. Phillips, D. G. Hinks, J. D. Jorgensen: *Phys. Rev. Lett.* **87**, 047001 (2001)
- 52.114 H.-J. Kim, W. N. Kang, E.-M. Choi, M.-S. Kim, K. H. P. Kim, S.-I. Lee: *Phys. Rev. Lett.* **87**, 087002 (2001)
- 52.115 W. N. Kang, H.-J. Kim, E.-M. Choi, K. H. P. Kim, S.-I. Lee: *Physica C* **378–381**, 1246 (2002)
- 52.116 S. L. Bud'ko, P. C. Canfield, V. G. Kogan: *Physica C* **382**, 85 (2002)
- 52.117 T. He, R. J. Cava, J. M. Rowell: *Appl. Phys. Lett.* **80**, 291 (2002)

Molecular Ele

53. Molecular Electronics

The prospects of using organic materials in electronics and optoelectronics applications have attracted scientists and technologists since the 1970s. This field has become known as **molecular electronics**. Some successes have already been achieved, for example the liquid-crystal display. Other products such as organic light-emitting displays, chemical sensors and plastic transistors are developing fast. There is also a keen interest in exploiting technologies at the molecular scale that might eventually replace silicon devices. This chapter provides some of the background physics and chemistry to the interdisciplinary subject of molecular electronics. A review of some of the possible application areas for organic materials is presented and some speculation is provided regarding future directions.

53.1	Electrically Conductive Organic Compounds	1220
53.1.1	Orbitals and Chemical Bonding	1220
53.1.2	Band Theory	1221
53.1.3	Electrical Conductivity	1222
53.2	Materials	1223
53.3	Plastic Electronics	1225
53.3.1	Diodes and Transistors	1225
53.3.2	Organic Light-Emitting Structures	1226
53.3.3	Photovoltaic Devices	1227
53.3.4	Chemical Sensors	1228
53.4	Molecular-Scale Electronics	1229
53.4.1	Moore's Laws	1229
53.4.2	Nanoscale Organic Films	1230
53.4.3	Patterning Technologies	1232
53.4.4	Molecular Device Architectures	1233
53.5	DNA Electronics	1235
53.6	Conclusions	1236
	References	1237

Molecular electronics is concerned with the exploitation of organic and biological materials in electronics and optoelectronics [53.1–3]. The subject, as it has matured over the last 30 years, can broadly be divided into two themes. The first area, *molecular materials for electronics*, has its origins in materials science and concerns the development of devices that utilise the unique macroscopic properties of organic compounds. The most successful commercial product to date is the liquid-crystal display. However, following many years of research, devices such as organic light-emitting displays, pyroelectric detectors for infrared imaging, and chemical and biochemical sensors are beginning to make a technological impact. The Nobel prize in Chemistry for 2000 was awarded to three scientists working in this area: Alan Heeger, Alan MacDiarmid and Hideki Shirakawa, who have made significant contributions to the development of electrically conductive polymers. More challenging is *molecular-scale electronics*. Here, the focus is on the behaviour of individual organic molecules or groups of molecules. Topics such as molecular switches [53.3–9], molecular memories [53.10–13], mo-

lecular rectifiers [53.14], negative differential-resistance junctions [53.15], deoxyribonucleic acid (DNA) electronics [53.16] and molecular manufacturing [53.17, 18] have all been described in the literature. It is much too early to say which, if any, of these could find their way into the commercial arena.

This chapter provides an introduction to the interdisciplinary world of molecular electronics. In the first instance, the physics background to semiconductive organic compounds is outlined. A review of the available materials is presented and some of the possible device applications are described. There are currently a limited number of ways in which organic molecules can be deposited and manipulated on surfaces to form solid films, which can then be used in device structures. The most popular methods are outlined, and methods to pattern the films are described. The prospects for molecular-scale electronics are contrasted with the progress of the inorganic semiconductor industry. Finally, a selection of the ongoing work on molecular-scale devices is described and some speculation about future developments is given.

53.1 Electrically Conductive Organic Compounds

Metallic and semiconductive behaviour is not restricted to inorganic materials. Figure 53.1 shows that the room-temperature conductivity values for organic polymers can extend over much of the spectrum of electrical conductivity, from insulating to semiconducting, and even metallic, behavior [53.19]. The physical explanation can be found in the nature of the chemical bonds that hold solids together.

53.1.1 Orbitals and Chemical Bonding

Carbon-based materials are unique in many respects. This is due to the many possible configurations of the electronic states of a carbon atom. Carbon has an atomic number of six and a valency of four. Its electron configuration is $1s^2, 2s^2, 2p^2$, i. e. the inner $1s$ shell is filled and the four valence electrons available for bonding are distributed two in the $2s$ orbital and two in the $2p$ orbitals. As the $2s$ orbital is spherically symmetrical, Fig. 53.2, it can form a bond in any direction. In contrast, the $2p$ orbitals are directed along mutually orthogonal axes and will tend to form bonds in these directions. When two or more of the valence electrons of carbon are involved in bonding with other atoms, the bonding can be explained by the construction of *hybrid* orbitals by mathematically combining the $2s$ and $2p$ orbitals. In the simplest

case, the carbon $2s$ orbital hybridises with a single p orbital. Taking the sum and difference of the two orbitals gives two sp *hybrids*; two p orbitals then remain. The sp orbitals are constructed from equal amounts of s and p orbitals; they are linear and 180° apart.

Other combinations of orbitals lead to different hybrids. For example, from the $2s$ orbital and two $2p$ orbitals (e.g. a $2p_x$ and a $2p_y$), three equivalent sp^2 hybrids may be constructed. Each orbital is 33.3% s and 66.7% p . The three hybrids lie in the xy plane (the same plane defined by the two p orbitals), directed 120° from each other, and the remaining p orbital is perpendicular to the sp^2 plane. Four sp^3 hybrids may be derived from an s orbital and three p orbitals. These are directed to the corners of a tetrahedron with an angle between the bonds of 109.5° ; each orbital is 25% s and 75% p .

A chemical bond can also be formed from a mixture of the above hybrid orbitals, e.g. it is possible to have a hybridised orbital that is 23% s and 77% p . Thus, sp , sp^2 and sp^3 hybrids must be considered as limiting cases. Electrons in s orbitals have a lower energy than electrons in p orbitals. Therefore, bonds with more s character tend to be stronger.

Carbon forms four bonds in most compounds, resulting from its four valence electrons. In ethane, C_2H_6 , the C–H bonds are all approximately $C(sp^3)-H(s)$ while the C–C bond is approximately $C(sp^3)-C(sp^3)$. However, in the ethylene molecule, C_2H_4 , each of the two carbons is attached to just three atoms while in acetylene, C_2H_2 , each carbon atom is only attached to one carbon. In these compounds, the two carbon atoms are bound together by double ($CH_2=CH_2$) or triple ($CH\equiv CH$) bonds, which involve sp^2 and sp hybrids, respectively. These bonds can be considered to have two distinct components. For example, in ethylene, two sp^2 hybrids on each carbon bond with the hydrogens. A third sp^2 hy-

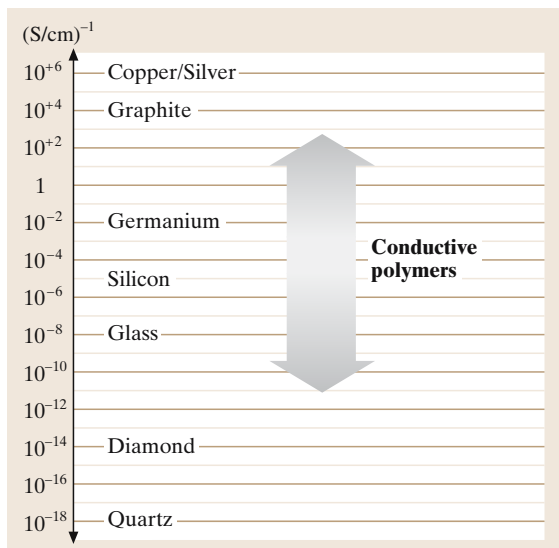


Fig. 53.1 Room-temperature conductivity values of conductive polymers compared to other materials. (After Roth [53.19], with permission © Wiley)

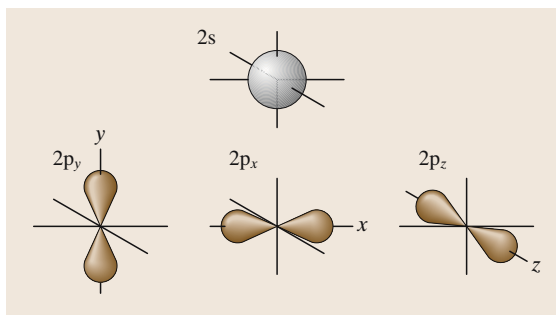


Fig. 53.2 Top: $2s$ orbital. Bottom: $2p_y$, $2p_x$ and $2p_z$ orbitals

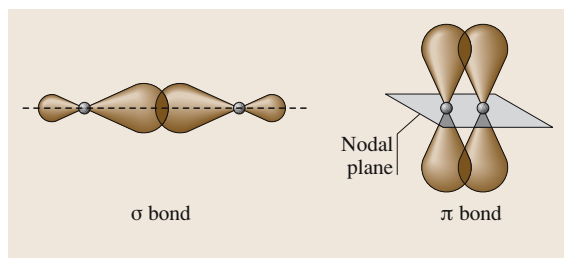


Fig. 53.3 Sigma (σ) and pi (π) bonding

brid on each carbon forms a $C(sp^2)-C(sp^2)$ single bond, leaving a p orbital 'left-over' on each carbon. This orbital lies perpendicular to the plane of the six atoms. The two p orbitals are parallel to each other and have regions of overlap above and below the molecular plane. This type of bond, consisting of two separate sausage-like electron clouds lying above and below the planes of the carbon nuclei, is called a pi (π) bond. In contrast, the bond formed by the head-on overlap of the two carbon sp^2 orbitals is known as a sigma (σ) bond and the electron cloud is densest at the midpoint between the carbon nuclei. The π bond has little electron density in the region between the carbon atoms and is much weaker than the σ bond. Figure 53.3 illustrates sigma and pi bonding. Carbon compounds that contain π bonds are said to be unsaturated, meaning that the carbon atoms, while having formed the requisite number of bonds, are not fully saturated in terms of their number of potential neighbours. Saturated carbon molecules, on the other hand, contain only single bonds.

53.1.2 Band Theory

The application of quantum mechanics to the electrons in the bonds of inorganic semiconductors led to the development of band theory, one of the great success stories of modern physics. Whenever two identical atoms are brought close together, the electron orbitals overlap and the energy level associated with each electron in the separated atoms is split into two new levels, with one above and one below the original level. A rigorous, quantum-mechanical description of the bonding reveals that the total number of orbitals must be conserved, i. e. the total number of molecular orbitals must be the same as the number of atomic orbitals that went into their formation. Consider, for example, the formation of a hydrogen molecule from two separated atoms. When the two atoms approach each other so that their 1s orbitals overlap, two new orbitals (σ bonds) are formed around the atoms, symmetric with respect to the interatomic axis. In one

orbital, the bonding orbital, the electron has a lower energy than in the isolated atomic orbital and in the other, the antibonding orbital, an electron has a higher energy. In the hydrogen molecule, the electron pair normally resides in the bonding orbital. The energy of the electrons in this orbital is lowered relative to that in the atomic orbitals, which is why the atoms remain bound together.

In an extended solid many atoms can interact and many similar splittings of energy levels occur. For a solid containing approximately 10^{26} atoms (Avogadro's number) each energy level splits, but the energies between these split levels are very small and continuous ranges or bands of energy are formed. Two such important bands are the valence band and the conduction band, analogous to the bonding and antibonding levels of the two-atom model. The energy gap, or band gap, between them is a forbidden energy range for electrons. Electrical conduction takes place by electrons moving under the influence of an applied electric field in the conduction band and/or holes moving in the valence band. Holes are really vacancies in a band but, for convenience, they may be regarded as positively charged carriers. For an electron, or hole, to gain energy from an applied electric field, and therefore for conductivity to occur, the charge carrier must be able to move into an unoccupied higher energy state. If the carrier cannot be accelerated by the field, then it cannot contribute to the electrical conduction. In a metal, the various energy bands overlap to provide a single energy band that is only partly full of electrons. An insulator has a full valence band and a relatively large energy separation (> 5 eV) to the higher conduction band. Most semiconductors possess the band structure of an insulator, but a forbidden energy gap of only 0.1–3 eV, so that carriers may be produced in the conduction and/or valence bands by optical or thermal means, or by doping with impurities.

An important feature of the band model is that the electrons are delocalised or spread over the lattice. The strength of the interaction between the overlapping orbitals determines the extent of delocalisation that is possible for a given system. For many polymeric organic materials, the molecular orbitals responsible for bonding the carbon atoms of the chain together are the sp^3 hybridised σ bonds, which do not give rise to extensive overlapping. The resulting band gap is large, as the electrons involved in the bonding are strongly localised on the carbon atoms and cannot contribute to the conduction process. This is why a simple saturated polymer such as polyethylene, $(CH_2)_n$, is an electrical insulator.

A significant increase in the degree of electron delocalisation may be found in unsaturated polymers, i. e.

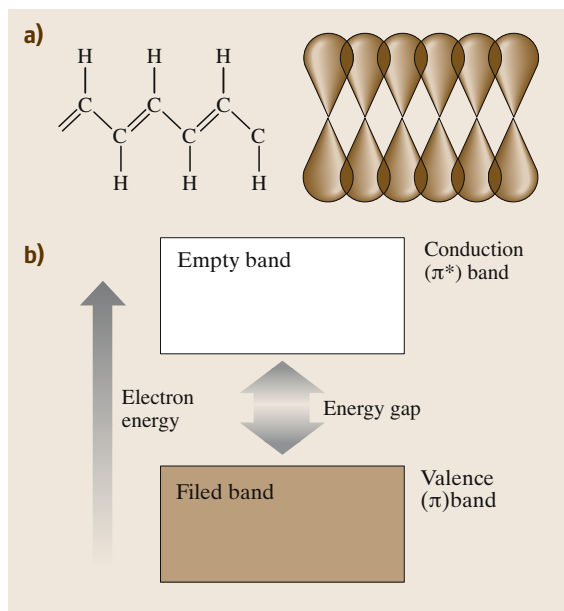


Fig. 53.4 (a) Chemical π bonding in polyacetylene. (b) Electronic band structure showing the normally empty π^* band (conduction band) and the normally filled π band (valence band)

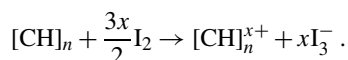
those containing double and triple carbon–carbon bonds. If each carbon atom along the chain has only one other atom, e.g. hydrogen, attached to it, the spare electron in a p_z -orbital of the carbon atom overlaps with those of carbon atoms on either side forming delocalised molecular orbitals of π -symmetry. For a simple lattice of length $L = Na$, where N is the total number of atoms and a is the spacing between them, it can be shown that the total number of electron states in the lowest energy band is equal to N . This result is true for every energy band in the system and applies to three-dimensional lattices. Allowing for the two spin orientations of an electron, the Pauli exclusion principle requires that there will be room for two electrons per cell of the lattice in an energy band. If each atom contributes one bonding electron, the valence band will be only half filled.

It might therefore be expected that a linear polymer backbone consisting of many strongly interacting coplanar p_z orbitals, each of which contributes one electron to the resultant continuous π electron system, would behave as a one-dimensional metal with a half-filled band. In chemical terms, this is a conjugated chain and may be represented by a system of alternating single and double bonds. It turns out that, for one-dimensional systems, such a chain can more efficiently lower its energy by in-

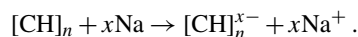
roducing bond alternation (alternating short and long bonds). This limits the extent of electronic delocalisation that can take place along the backbone. The effect is to open an energy gap in the electronic structure of the polymer. All conjugated polymers are semiconductors, with band gaps more than about 1.5 eV, rather than metals (N.B. the band gap in silicon is 1.1 eV at room temperature). Figure 53.4 shows the chemical π bonding and the electronic band structure of polyacetylene.

53.1.3 Electrical Conductivity

In polyacetylene, the valence band, or π band, is normally filled with electrons while the conduction band, or π^* band is normally empty. Like silicon, the conductivity of polyacetylene can be changed by the addition of impurity atoms. However, the term *doping* is a misnomer as it tends to imply the use of minute quantities, parts per million or less, of impurities introduced into a crystal lattice. In the case of conductive polymers, typically 1% to 50% by weight of chemically oxidising (electron withdrawing) or reducing (electron donating) agents are used to alter physically the number of π -electrons on the polymer backbone, leaving oppositely charged counterions alongside the polymer chain. These processes are redox chemistry. For example, the halogen doping process that transforms polyacetylene to a good conductor is oxidation (or p-doping).



Reductive doping (n-doping) is also possible, e.g. using an alkali metal.



In both cases, the doped polymer is a salt. The counterions, I_3^- or Na^+ , are fixed in position while the charges on the polymer backbone are mobile and contribute to the conductivity. The doping effect can be achieved because a π electron can be removed (or added) without destroying the σ backbone of the polymer so that the charged polymer remains intact. The increase in conductivity can be as much as eleven orders of magnitude.

The electrical properties of semiconductive organic polymers are not directly comparable to those of silicon. An important material parameter is the mobility of the charge carriers μ . This determines the (additional) velocity that a charge carrier (an electron or hole) acquires because of an applied electric field, and is defined by

$$\mu = \frac{v_d}{E},$$

Table 53.1 Room-temperature field-effect carrier mobilities for field-effect transistors based on organic semiconductors [53.20,21]. The electron mobilities in single-crystal silicon and gallium arsenide are also given

Material	Mobility (cm^2/Vs)
Si single crystal (electrons)	1500
GaAs single crystal (electrons)	8500
Polythiophene	10^{-5}
Polyacetylene	10^{-4}
Phthalocyanine	$10^{-4} - 10^{-2}$
Thiophene oligomers	$10^{-4} - 10^{-1}$
Organometallic dmit complex	0.2
Pentacene	$10^{-3} - 1$
C_{60}	0.3

where v_d is the drift velocity of the carrier and E is the electric field. The mobility may be further related to the electrical conductivity σ by the expression

$$\sigma = |q|n\mu,$$

where n is the density of charge carriers and $|q|$ is the magnitude of their charge (charge on an electron $= 1.6 \times 10^{-19} \text{ C}$). The carrier mobility provides an indication of how quickly the carriers react to the field (i.e.

the frequency response of the material). The greater the degree of electron delocalisation, the larger the width of the bands (in energy terms) and the higher the mobility of the carriers within the band. For inorganic semiconductors such as silicon or gallium arsenide, the three-dimensional crystallographic structure provides for extensive carrier delocalisation throughout the solid, resulting in a relatively high mobility μ .

Electrical conduction in polymers not only requires carrier transport along the polymer chains but some kind of transfer, or *hopping*, between these chains, which tend to lie tangled up like a plate of spaghetti. The charge-carrier mobilities in organic polymers are therefore quite low, making it difficult to produce very high-speed electronic computational devices that are competitive with those based on silicon and gallium arsenide. However, some improvement in the carrier mobility can be achieved by both increasing the degree of order of the polymer chains and by improving the purity of the material. Table 53.1 contrasts the room-temperature carrier mobility values for Si, GaAs and a number of different conductive organic compounds [53.20,21]. Although the values are quite low for the organic materials, other features make them attractive for certain types of electronic device, as indicated in the later sections.

53.2 Materials

Many conductive polymers have been synthesised to provide certain electronic features (e.g. band gap, electron affinity). The monomer repeat units are often based on five- or six-membered (benzene) carbon ring systems, including polypyrrole, polythiophene (and various other polythiophene derivatives), polyphenylenevinylene and polyaniline [53.22]. The chemical structure of some of these materials is shown in Fig. 53.5. Such polymers generally show lower electrical conductivity than polyacetylene, however they can have the advantage of high stability and processability.

Conductive polymers represent only one category of organic electrical conductors. Another important class are the charge-transfer compounds [53.23, 24]. These are formed from a variety of molecules, primarily aromatic compounds (i.e. based on benzene) which can behave as electron donors (d) and electron acceptors (a). Complete transfer of an electron from a donor to an acceptor molecule results in a system that is electrically insulating (e.g. the transfer of a valence electron in a Na

atom to a Cl atom, forming the compound NaCl). However, if the ratio of the number of donor molecules to the number of acceptor molecules differs from 1 : 1, e.g. the stoichiometry is 1 : 2 or 2 : 3, or if there is incomplete transfer of an electron from a donor to an acceptor (say,

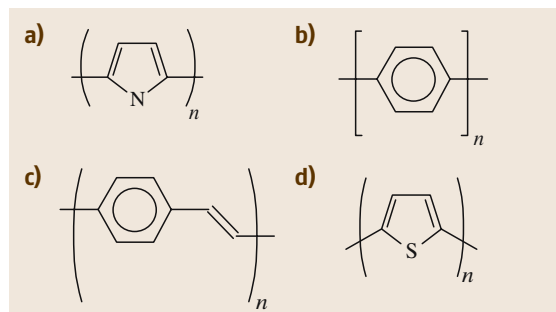


Fig. 53.5a–d Chemical structures of conductive polymers: (a) polypyrrole, (b) polyparaphenylene, (c) polyphenylenevinylene, and (d) polythiophene

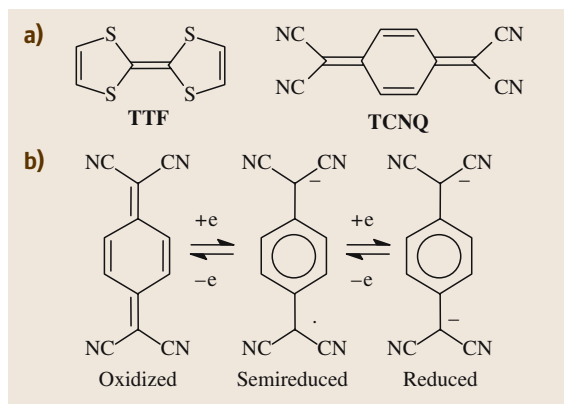
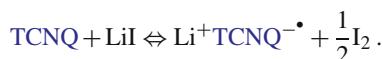


Fig. 53.6 (a) Chemical structures of the charge-transfer compounds tetrathiafulvalene (TTF) and tetracyanoquinodimethane (TCNQ). (b) Oxidation and reduction of TCNQ

six electrons in every ten are transferred), then partially filled electron energy bands can be formed and electron conduction is possible.

Charge-transfer interactions are strongest between donor molecules of low ionisation potential and acceptor molecules of high electron affinity, provided that the donor and acceptor molecules have similar symmetry and are able to approach closely [53.24]. Well-known donor and acceptor molecules are tetrathiafulvalene (TTF) and tetracyanoquinodimethane (TCNQ), Fig. 53.6a. The latter compound is a very strong acceptor forming first the radical anion and then the dianion, Fig. 53.6b. The stability of the semireduced radical ion, $\text{TCNQ}^{\cdot-}$, with respect to the neutral molecule mainly arises from the change from the relatively unstable quinoid structure to the aromatic one, allowing extensive delocalisation of the π electrons over the carbon skeleton. As a consequence, TCNQ not only forms typical charge-transfer complexes but is also able to form true radical-ion salts, incurring complete one-electron transfer. Thus, on addition of lithium iodide to a solution of TCNQ, the simple lithium TCNQ salt is formed:



Following removal of the free iodine precipitate, the TCNQ salt may be crystallised. The crystals show an electronic conductivity of about 10^{-5} S/cm.

A 1 : 1 TCNQ : TTF salt exhibits a high room-temperature conductivity (5×10^2 S/cm) and metallic behaviour is observed as the temperature is reduced to 54 K. The molecules in such compounds are arranged in segregated stacks, in which the donors and accep-

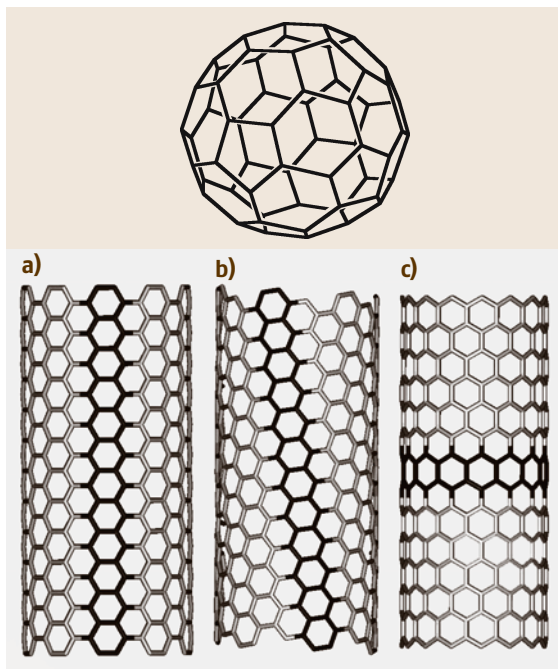


Fig. 53.7a–c Chemical structure of C₆₀ (top) and three classes of single-wall carbon nanotube (SWNT) (a) (10,10) armchair SWNT, (b) (12,7) chiral SWNT and (c) (15,0) zigzag SWNT

tors form separate donor stacks (ddddd...) and acceptor stacks (aaaaa...). The molecules are stacked in such a way that the π bonds on successive molecules can overlap to form bands. This overlap is different from the p-orbital overlap forming the π bands in conjugated polymers. The crystalline packing in these, and other, charge-transfer compounds generally leads to carrier mobility values that are higher than for semiconductive polymers.

Other electroactive compounds that may find application in molecular electronics are based on forms of pure carbon. Graphite consists of vast carbon sheets stacked one on top of another like a sheaf of papers. In pure graphite, these layers are about 0.335 nm apart, but they can be separated further by intercalating various molecules. The bonding between the carbon atoms in the planes is mainly sp^2 hybridisations consisting of a network of single and double bonds. Weak interactions between the delocalised electron orbitals hold adjacent sheets together. The delocalised electron system in the planes results in semiconductive electrical behaviour.

Under certain conditions, carbon forms regular clusters of 60, 70, 84 etc. atoms [53.25, 26]. A C₆₀ cluster,

shown in Fig. 53.7, is composed of 20 hexagons and 12 pentagons and resembles a football. The diameter of the ball is about 1 nm. As with graphite, each carbon atom in C_{60} is bonded to three other carbon atoms. Thus C_{60} can be considered as a rolled up layer of a single graphite sheet. The term buckminsterfullerene was given originally to the C_{60} molecule because of the resemblance to the geodesic domes designed and built by Richard Buckminster Fuller. However, this term (or fullerene or buckyball) is used quite generally to describe C_{60} and related compounds. For example, a molecule with the formula C_{70} can be formed by inserting an extra ring of hexagons around the equator of the sphere, producing an elongated shell more like a rugby ball.

In addition to the spherical-shaped fullerenes, it is possible to synthesise tubular variations – carbon nanotubes [53.26, 27]. Such tubes are comprised of graphite-like sheets curled into a cylinder. Each tube may contain several cylinders nested inside each other. The tubes are capped at the end by cones or faceted hemispheres. Because of their very small diameters (down to around 0.7 nm), carbon nanotubes are prototype one-dimensional nanostructures. An important feature of a carbon nanotube is the orientation of the six-membered carbon ring in the honeycomb lattice relative to the axis of the nanotube. Three examples of single-wall carbon nanotubes (SWCNs) are shown in Fig. 53.7. The primary classification of a carbon nanotube is as either being chiral or achiral. An achiral nanotube is one whose mirror image has an identical structure to the original. There are only two cases of achiral nanotubes: armchair and zigzag (these names arise from the shape of the cross-sectional ring). Chiral nanotubes exhibit a spiral symmetry whose

mirror image cannot be superimposed on the original structure. Carbon nanotubes are characterised by the chiral index (n, m) where the integers n and m specify each carbon nanotube uniquely [53.27]. An armchair nanotube corresponds to the special case $n = m$, while for a zigzag nanotube $m = 0$. All other (n, m) indexes correspond to chiral nanotubes. The electronic structure of a SWCN is either metallic or semiconducting, depending on its diameter and chirality.

At low temperature, a single-wall carbon nanotube is a quantum wire in which the electrons in the wire move without being scattered. Resistance measurements for various nanotube samples show that there are metallic and semiconducting nanotubes [53.27]. Carbon nanotubes can also be doped either by electron donors or electron acceptors [53.28]. After reaction with the host materials, the dopants are intercalated in the intershell spaces of the multiwalled nanotubes, and, in the case of single-walled nanotubes, either in between the individual tubes or inside the tubes.

The above confirms carbon's uniqueness as an electronic material. It can be a good conductor in the form of graphite, an insulator in the form of diamond, or a flexible polymer (conductive or insulating) when reacted with hydrogen and other species. Carbon differs from other group IV elements, such as Si and Ge, which exhibit sp^3 hybridisation. Carbon does not have any inner atomic orbitals except for the spherical 1s orbital, and the absence of nearby inner orbitals facilitates hybridisations involving only the valence (outer) s and p orbitals. The fact that sp and sp^2 hybridisations do not readily occur in Si and Ge might be related to the absence of *organic materials* made from these elements.

53.3 Plastic Electronics

53.3.1 Diodes and Transistors

Since the discovery of semiconducting behaviour in organic materials, there has been a considerable research effort aimed at exploiting these properties in electronic and optoelectronic devices. The term *plastic electronics* refers to electronic devices incorporating polymeric organic compounds (although this term is often used more widely to include devices incorporating other semiconducting organic materials). Organic semiconductors can have significant advantages over their inorganic counterparts. For example, thin layers

of polymers can easily be made by low-cost methods such as spin coating. High-temperature deposition from vapour reactants is generally needed for inorganic semiconductors. Synthetic organic chemistry also offers the possibility of designing new materials with different band gaps. As noted in Sect. 53.1.3, the mobilities of the charge carriers in organic field-effect transistors are low. Nevertheless, the simple fabrication techniques for polymers have attracted several companies to work on polymer transistor applications such as data storage and thin-film device arrays to address liquid-crystal displays [53.20, 29–31].

Semiconducting organic compounds have been used in a similar fashion to inorganic semiconductors (e.g. Si, GaAs) in metal/semiconductor/metal structures. A diode, or rectifying device, can be made by sandwiching a semiconductor between metals of different work functions. In the ideal case, an n-type semiconductor should make an Ohmic contact to a low-work-function metal and a rectifying Schottky barrier to a high-work-function metal [53.32]. One example is that of a semiconductive organic film sandwiched between aluminium and indium-tin-oxide electrodes [53.33]. This device also exhibits photovoltaic behaviour.

Organic materials have been used as the semiconducting layer in field-effect transistor (FET) devices [53.20, 34–36]. These are three-terminal structures: a voltage applied to a metallic gate affects an electric current flowing between the source and drain electrodes. For transistor operation, charge must be injected easily from the source electrode into the organic semiconductor and the carrier mobility should be high enough to allow useful quantities of source–drain current to flow. The organic semiconductor and other materials with which it is in contact must also withstand the operating conditions without thermal, electrochemical or photochemical degradation. Two performance parameters to be optimised in organic field-effect transistors are the field-effect mobility and the on/off ratio [53.35].

The operating characteristics of organic transistors and integrated circuits have improved markedly over recent years. This has been brought about by both improvements in the material synthesis and in the thin-film processing techniques [53.37–45]. State-of-the-art organic FETs possess characteristics similar to those of devices prepared from hydrogenated amorphous silicon, with mobilities around $1 \text{ cm}^2/\text{Vs}$ and on/off ratios greater than 10^6 . A number of groups have also demonstrated transistor devices incorporating carbon nanotubes [53.46–48]. However, some key issues need to be addressed before nanotubes can be exploited fully in such applications. These include the reproducible fabrication of low-resistance electrical contacts and the accurate control of nanotube growth parameters [53.49]. The nanotube transistor devices realised experimentally have typical dimensions in the micron range. The real promise of carbon nanotube devices, however, lies in the possibility of nanoscale devices.

Thin-film transistors based on organic semiconductors are likely to form key components of plastic circuitry for use as display drivers in portable computers and pagers, and as memory elements in transaction cards and identification tags.

53.3.2 Organic Light-Emitting Structures

Reports of light emission from organic materials on the application of an electric field (electroluminescence) have been around for many years. However, there has been an upsurge in interest following the initial report of organic light-emitting devices (OLEDs) incorporating the conjugated polymer polyphenylenevinylene (PPV, Fig. 53.5) [53.50]. The simplest OLED is an electroluminescent compound sandwiched between metals of high and low work function, as depicted in Fig. 53.8. The anode electrode is normally indium tin oxide (ITO) as this material is semitransparent, allowing the light out of the device. On application of a voltage, electrons are injected from the low work function electrode into the lowest unoccupied molecular orbital (LUMO) level (conduction or π^* band in the case of an organic compound possessing a delocalised electron system) of the organic compound and holes from the high-work-function electrode into the highest occupied molecular orbital (HOMO) level (valence or π band in the case of an organic compound possessing a delocalised electron system). The recombination of these oppositely charged carriers then results in the emission of light. Work is focused on the use of low-molecular-weight organic molecules and polymers and there is considerable industrial interest in the application of such materials to various display technologies [53.51–54]. It is estimated that the global market for OLED displays will

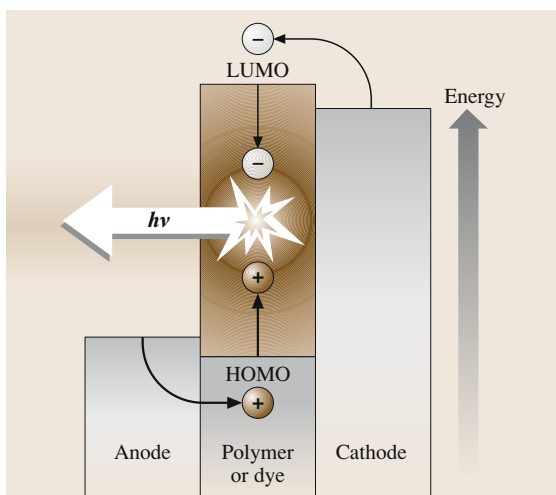


Fig. 53.8 Schematic energy band structure of an organic light-emitting device (OLED). The recombination of electrons and holes results in the emission of light of frequency ν and energy $h\nu$

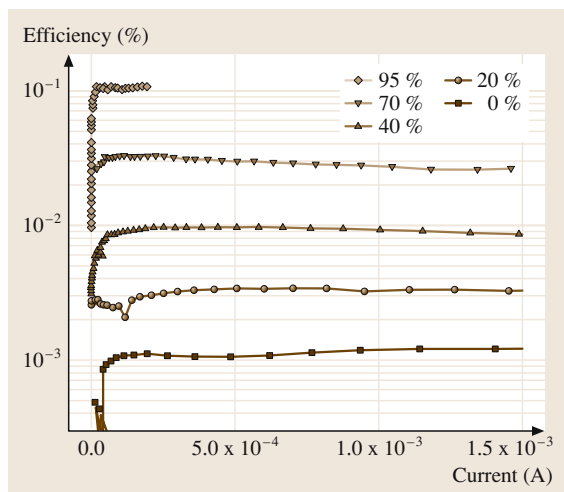


Fig. 53.9 External quantum efficiency versus current for OLEDs fabricated with MEH-PPV/DFD blended layers using blends with various concentrations (by weight) of DFD. (After *Ahn et al.* [53.56], with permission, © AIP)

increase from \$112 million in 2003 to almost \$2.3 billion by 2008 [53.55]. Organic light-emitting devices offer a number of advantages over liquid-crystal displays (LCDs) and the other technologies that currently dominate the flat-panel display market: OLEDs do not require backlighting and they can be made thinner and lighter. Furthermore, OLED displays provide higher contrast and truer colours, higher brightness, wider viewing angles, better temperature stability and faster response times than LCDs [53.54].

Many techniques have been used in attempts to optimise the performance of OLEDs. For example, a thin inorganic insulating layer such as LiF, or an organic monolayer may be inserted between the cathode and the emissive material [53.57, 58]. Electron- and hole-transporting layers can also be introduced between the cathode and the emitting layer and between the anode and the emitting layer, respectively, to improve and balance the injection of charge carriers [53.51]. As an alternative, OLEDs based on blended single layers of emissive and charge-transport materials may be fabricated [53.56, 59–61]. Such devices generally possess lower efficiencies than multilayer structures but they have the considerable advantage of ease of manufacture. Figure 53.9 shows how the external quantum efficiency of a blended layer OLED based on a polymer – poly[2,–(2-ethylhexyloxy)–5-methoxy–1,4-phenylenevinylene] (MEH-PPV) – and an electron-

Table 53.2 Performance of polymer-based organic light-emitting devices. (Data from Cambridge Display Technology [53.54])

Colour	Voltage (V)	Efficiency at 100 cd/m ² (lm/W)	Measured lifetime at room temperature at fixed luminance (h) at (cd/m ²)
Red	3.6	1.5	1790 at 2000
Green	2.6	9.4	2867 at 2000
Blue	5.5	3.9	1426 at 720
Yellow	3.3	6.5	2420 at 4000

^a All data taken using common cathode and may include an interlayer

transport compound – 2,7-bis[2–(4-tert-butylphenyl)–1,3,5-oxadiazol-5-yl]–9,9-dihexylfluorene (DFD) – varies as the ratio MEH-PPV : DFD is changed in a spin-coated film [53.56]. The device efficiency increases with the amount of DFD so that the OLED fabricated with a 95% blend is about 100 times more efficient than that of pure MEH-PPV. For all the devices, the electroluminescence originates from the MEH-PPV material, indicating that the energy, or charge, transfer between the electron-transport compound and the polymer is very efficient.

When an electron and a hole recombine to form an excited molecular state in an organic material, the spins of the electrons in the excited and ground levels can either point in the same direction (triplet state) or in the opposite directions (singlet state). For quantum-mechanical reasons, 75% of recombination events are associated with triplet states which, in most cases, do not emit photons when they decay to the ground state. Hence, the production of emission from the triplet state of organic materials is a further means to improve the device efficiency [53.52, 62].

Table 53.2 shows some important parameters of polymer-based OLEDs with different colour outputs [53.54]. There is also a keen interest in developing white-light organic displays. For example, a large-area white-light-emitting OLED could provide a solid-state light source that might compete with conventional lighting technologies. Different methods of making an intrinsically white-emitting OLED by blending emissive species, either in single or multiple layers have been demonstrated [53.63–65]. Alternatively, a blue OLED can be used with one or more down-conversion layers [53.66].

53.3.3 Photovoltaic Devices

Concerns over global climate change, local air pollution and resource depletion are making photovoltaics (PVs) an increasingly attractive method of energy supply. The current technology is based on single-crystal silicon solar cells. These have developed since the 1940s and now possess conversion efficiencies of around 15% for commercial devices (although figures of around 25% are reported in the laboratory). However, the technology is more expensive than conventional power generation and there is much research on alternative materials. Photovoltaics using organic compounds, such as polymers or dyes, offer the possibility of large-scale manufacture at low temperature coupled with low cost. Until the end of the 20th century little progress had been made and energy conversion efficiencies of up to only about 1% were achieved. However, the availability of new conductive organic materials and different PV designs have significantly improved on this figure. To 2004, several laboratories have reported conversion efficiencies of 4–5%, with lifetimes of around 10⁴ hours [53.67].

An organic solar cell device is very similar in structure to the OLED described in the previous section. If the incoming photons have energy greater than the band gap of the polymer (or greater than the HOMO–LUMO separation in the case of organic molecular materials) then the light will be absorbed, creating electrons and holes. In an inorganic photovoltaic cell, these electrons and holes would be generated within, or close to, a depletion region in the semiconductor and they would be free to migrate to opposite electrodes, where they can do useful work in an external electrical load. However, in the organic material the electrons and holes are bound together in excitons. An immediate problem in organic PV cells is to split these excitons. This can be conveniently done at an interface, the simplest being the junction between the electrodes and the organic material. Under open-circuit conditions, holes are collected at the high-work-function electrode (e.g. ITO) and electrons at the low-work-function electrode (e.g. Al). The open-circuit output voltage of the PV device depends on the work-function difference between the electrodes. Improvements in the efficiency of the exciton-splitting process can be achieved using organic compounds incorporating electron-donating and electron-accepting species. By creating interfaces of differing electron affinities, it is possible to enhance the probability of electron transfer between the molecules.

An alternative approach to organic PVs exploits a dye-sensitised solar cell, or Grätzel cell [53.68]. Here,

the incoming photons are absorbed by molecules of a dye on a semiconductor surface with subsequent energy and electron transfer to the semiconductor. An electron is returned to the oxidised dye via an electrolyte. The efficiency of such devices can approach 10% [53.69].

53.3.4 Chemical Sensors

The development of effective devices for the identification and quantification of chemical and biochemical substances for process control and environmental monitoring is a growing need [53.70–72]. Many sensors do not possess the specifications to conform to existing or forthcoming legislation; some systems are too bulky/expensive for use in the field. Inorganic materials such as the oxides of tin and zinc have traditionally been favoured as the sensing element [53.73]. However, one disadvantage of sensors based on metallic oxides is that they usually have to be operated at elevated temperatures, limiting some applications. As an alternative, there has been considerable interest in trying to exploit the properties of organic materials. Many such substances, in particular phthalocyanine derivatives, are known to exhibit a high sensitivity to gases [53.74]. Lessons can also be taken from the biological world; one household carbon-monoxide detector is designed to simulate the reaction between CO and haemoglobin. A significant advantage of organic compounds is that their sensitivity and selectivity can be tailored to a particular application by modifications to their chemical structure. Moreover, thin-film technologies, such as self-assembly or layer-by-layer electrostatic deposition, enable ultrathin layers of organic materials to be engineered at the molecular level [53.75].

There are many physical principles upon which sensing systems might be based; changes in electrical resistance (chemiresistors), refractive index (fibre-optic sensors) and mass (quartz microbalance) have all been exploited in chemical sensing. The main challenges in the development of new sensors are in the production of cheap, reproducible and reliable devices with adequate sensitivities and selectivities.

A simple chemiresistor sensor exploits the resistance change of a thin layer of a gas-sensitive material. For example, the conductivity of phthalocyanine thin films can be changed in the presence of oxidising or reducing gases [53.76]. This effect is analogous to the doping of an inorganic semiconductor, such as silicon, with acceptor or donor impurities (Sect. 53.1.3). A problem associated with these chemiresistor devices is that the current outputs are low (typically picoamperes) requiring elaborate

detection electronics and careful shielding and guarding of components. This difficulty may be overcome by incorporating the organic sensing layer into a silicon field-effect transistor. A schematic diagram of such a structure is shown in Fig. 53.10 [53.77, 78]. Changes in the sheet resistance of the sensing layer are reflected in the variation in both the amplitude and phase of the transfer function of the device. These devices have been shown to respond well to low (vapour parts per million) concentrations of NO_2 [53.79]. It is also possible to use a chemically sensitive organic film as the semiconductive layer in a diode or a transistor. In these cases the device must be fabricated so that the gas can interact readily with the organic material.

While many sensing devices show adequate sensitivities, the selectivity can be poor. For example, a semiconductive polymer may show a similar change in electrical resistance to a range of oxidising (reducing) gases. To get around this difficulty, one approach that is being embraced enthusiastically by researchers is to use an array of sensing elements, rather than a single device. This is the method favoured by nature. The human olfactory system has many receptor cells (sensors), which are individually nonspecific; signals from these are fed to the brain via a network of primary and secondary neurons for processing. It is generally believed

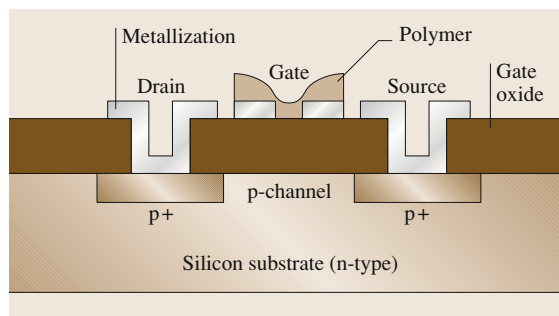


Fig. 53.10 Structure of hybrid silicon/organic field-effect transistor for chemical sensing. (After Barker et al. [53.78])

that the selectivity of the olfactory system is a result of a high degree of parallel processing in the neural architecture. Artificial neural networks can emulate the connectivity of the olfactory neurons. The electronic nose is an attempt to mimic the human olfactory system and there are now several companies marketing such equipment [53.80]. Individual sensors can be based on polymer films. Each element is treated in a slightly different way during deposition so that it responds uniquely on exposure to a particular gas or vapour. The pattern of resistance changes in the sensor array can then be used to fingerprint the vapour.

53.4 Molecular-Scale Electronics

53.4.1 Moore's Laws

The second strand to molecular electronics (molecular-scale electronics) recognises the spectacular size reduction in the individual processing elements in integrated circuits over recent years. The first microprocessor chip manufactured in 1972 by Intel (8008) had a clock speed of 200 kHz and contained 3500 transistors. There are 55 million transistors on the Pentium® 4 chip (November, 2002), fabricated using 0.13- μm process technology, and operating at a clock speed of 3 GHz.

Moore's law (or Moore's first law) states that the functions per chip double every 1.5 years. This will probably describe developments over at least the next decade. The semiconductor industries have produced an international technology roadmap for the future of complementary metal oxide semiconductor (CMOS) technology [53.81]. Figure 53.11 shows the anticipated growth in the density of the transistors in both the microprocessor unit (MPU) and the dynamic random-access memory (DRAM) of a CMOS chip over the next decade.

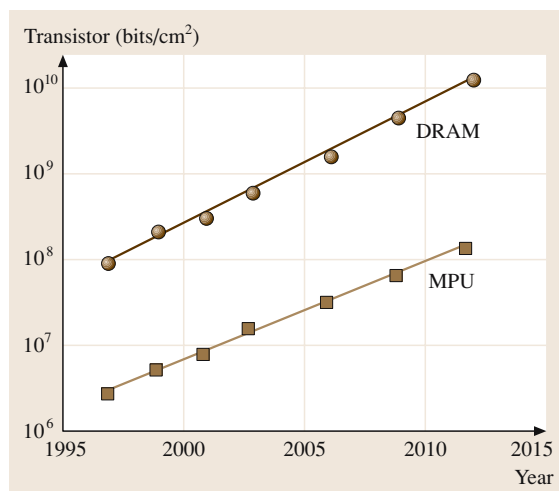


Fig. 53.11 Predicted feature size in CMOS devices. (After [53.81])

Table 53.3 Information content for various applications. (After Tour [53.3])

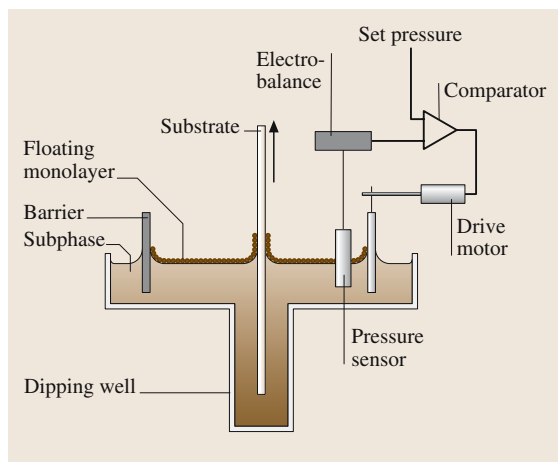
Application	Typical information content (bytes)
Color photograph	10^5
Average book	10^6
Desktop computer	10^8
Genetic code	10^{10}
Human brain	10^{13}
Library of congress	10^{15}

The prediction is for a 30-nm minimum feature size (gate length) for the MPU and 10^{10} transistors per cm^2 in the case of the memory by the year 2012. These figures are regularly updated. This device density is remarkable and provides the means to store significant amounts of data, Table 53.3 [53.3]. However, molecular-scale electronics has the potential for further increases in device density. For example, using $\sim 1\text{--}3$ nm organic molecules as the processing elements, $10^{13}\text{--}10^{14}$ ‘devices’ could be fitted into 1 cm^2 [53.3].

There are, however, a number of technological problems issues that will need to be overcome for the predictions for the CMOS-based roadmap to be realised [53.82]. Not least, are the materials limitations of the silicon/silicon dioxide system. For example, charge leakage becomes a problem when the insulating silicon dioxide layers are thinned to a few nm. Heat dissipation is another key factor. A Pentium® 4 chip, with $10^7\text{--}10^8$ transistors operating at the current nanosecond rate can emit 50–60 W of heat. Financial issues are also problematic as ever more complex integrated circuits are produced. Intel’s Fab 22, a chip-fabrication facility (FAB) which opened in Chandler, Arizona in October 2001, cost \$2 billion to construct and equip. The cost of building a fab is projected to rise to \$15–30 billion by 2010 and could be as much as \$200 billion by 2015 [53.83]. This significant increase in cost (Moore’s second law) is due to the extremely sophisticated tools that will be needed to form the increasingly small features of the devices. Molecular-scale technology will, of course, also need to address such problems – plus many more.

53.4.2 Nanoscale Organic Films

Before organic molecules can be exploited in device architectures (i.e. those familiar from Si-based microelectronics) ways must be found to deposit them onto surfaces. Well-established methods of organic film de-

**Fig. 53.12** System for the deposition of Langmuir–Blodgett films

position include electrodeposition, thermal evaporation and spinning [53.84]. The Langmuir–Blodgett (LB) technique, self-assembly and layer-by-layer electrostatic deposition are further means for producing layers of organic materials. These allow ultrathin-film assemblies of organic molecules to be engineered at the molecular level and are of particular relevance to molecular-scale electronics [53.85, 86].

Langmuir–Blodgett films are prepared by first depositing a small quantity of an amphiphilic compound (i.e. one containing both polar and nonpolar groups) dissolved in a volatile solvent onto the surface of purified water [53.86, 87]. The classical materials are long-chain fatty acids, such as *n*-octadecanoic acid (stearic acid). When the solvent has evaporated, the organic molecules may be organised into a floating two-dimensional *crystal* by compression on the water surface. As the area available to the organic molecules is reduced, the floating film will undergo several phase transformations. These are, to a first approximation, analogous to three-dimensional gas, liquid and solid phases. The phase changes may readily be identified by monitoring the surface pressure as a function of the area occupied by the molecules in the film. This is the two-dimensional equivalent to the pressure-versus-volume isotherm for a gas. In the *gaseous* state the molecules are far enough apart on the water surface that they exert little force on one another. As the surface area of the monolayer is reduced, the hydrocarbon chains will begin to interact. The *liquid* state that is formed is generally called the expanded monolayer phase. The hydrocarbon chains of the molecules in such a film are in a random, rather than regular orienta-

tion, with their polar groups in contact with the subphase. As the molecular area is progressively reduced, condensed phases may appear. There may be more than one of these and the emergence of each condensed phase can be accompanied by constant-pressure regions of the isotherm, as observed in the cases of a gas condensing to a liquid and a liquid solidifying. In the condensed monolayer states, the molecules are closely packed and are oriented with their hydrocarbon chain pointing away from the water surface. The area per molecule in such a state will be similar to the cross-sectional area of the hydrocarbon chain, i. e. $\approx 0.19 \text{ nm}^2/\text{molecule}$.

If the surface pressure is held constant in one of the condensed phases, then the film may be transferred from the water surface onto a suitable solid substrate simply by raising and lowering the latter through the monolayer–air interface. Figure 53.12 shows a schematic diagram of the equipment required for LB film deposition. In this technique, introduced by Langmuir and Blodgett [53.86] the floating condensed monolayer is transferred, like a carpet, as the substrate is raised and/or lowered through the air/monolayer interface. A number of different LB deposition *modes* are possible. The most commonly encountered situation is Y-type deposition, which refers to monolayer transfer on both the upward and downward movements of the substrate. Instances in which the floating monolayer is only transferred to the substrate as it is being inserted into the subphase, or only as it is being removed, are also observed. These deposition modes are called X-type and Z-type deposition, respectively.

Self-assembly is a much simpler process than that of LB deposition. Monomolecular layers are formed by the immersion of an appropriate substrate into a solution of the organic material [53.85]. The best known examples of self-assembled systems are organosilicon on hydroxylated surfaces (SiO_2 , Al_2O_3 , glass etc.) and alkanethiols on gold, silver and copper. However, other combinations include: dialkyl sulphides on gold; dialkyl disulphides on gold; alcohols and amines on platinum; and carboxylic acids on aluminium oxide and silver. The self-assembly process is driven by the interactions between the head group of the self-assembling molecule and the substrate, resulting in a strong chemical bond between the head group and a specific surface site, e.g. a covalent Si–O bond for alkyltrichlorosilanes on hydroxylated surfaces.

The combination of the self-assembly process with molecular recognition offers a powerful route to the development of nanoscale systems that may have technological applications as chemical sensing or switching

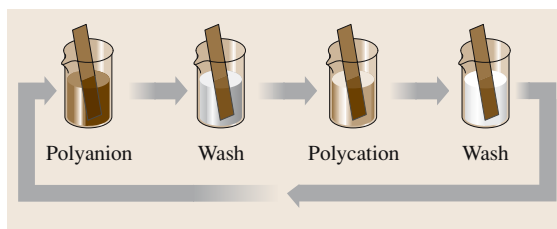


Fig. 53.13 Deposition of layer-by-layer polyelectrolyte films by the sequential immersion of a solid substrate in solutions of the polyanions and polycations

devices. For instance, the complexation of a neutral or ionic guest at one site in a molecule may induce a change in the optical or redox properties of the system. Monolayers containing both an electroactive TTF unit and a metal-binding macrocycle assembled onto platinum have been shown to exhibit electrochemical recognition to Ag^+ ions [53.88].

The self-assembly process, as described above, is usually restricted to the deposition of a single molecular layer on a solid substrate. However, chemical means can be exploited to build up multilayer organic films. A method pioneered by Sagiv is based on the successive absorption and reaction of appropriate molecules [53.89, 90]. The head groups react with the substrate to give a permanent chemical attachment and each subsequent layer is chemically attached to the one before in a very similar way to that used in systems for supported synthesis of proteins.

Another technique for building up thin films of organic molecules is driven by the ionic attraction between opposite charges in two different polyelectrolytes, the so-called *layer-by-layer* assembly technique [53.91–93]. A solid substrate with a positively charged planar surface is immersed in a solution containing an anionic polyelectrolyte and a monolayer of polyanion is adsorbed, Fig. 53.13. Since the adsorption is carried out at relatively high concentrations of the polyelectrolyte, most of the ionic groups remain exposed to the interface with the solution and thus the surface charge is reversed. After rinsing in pure water, the substrate is immersed in a solution containing the cationic polyelectrolyte. Again, a monolayer is adsorbed but now the original surface charge is restored, resulting in the formation of a multilayer assembly of both polymers. It is possible to use a sensitive optical technique, such as surface plasmon resonance to monitor, in situ, the growth of such electrostatically assembled films [53.94]. The layer-by-layer method has been used to build up layers of conductive poly-

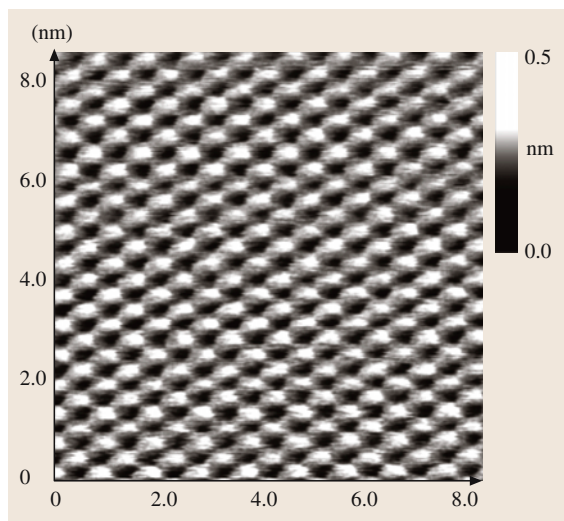


Fig. 53.14 Atomic force micrograph of a 12-layer, *n*-eicosanoic-acid Langmuir–Blodgett film deposited onto silicon. (After, in part, with permission from *Evanson et al.* [53.98]. © (1996) ACS)

mers, e.g. partially doped polyaniline and a polystyrene polyanion [53.95]. Biocompatible surfaces consisting of alternate layers of charged polysaccharides and oppositely charged synthetic polymers can also be deposited in this way [53.96]. A related, but alternative, approach uses layer-by-layer adsorption driven by hydrogen-bonding interactions [53.97].

The organisation of the organic molecules in multilayer assemblies may be investigated by a number of analytical techniques including X-ray and neutron reflection, electron diffraction and infrared spectroscopy [53.86]. Figure 53.14 shows an example of an atomic force micrograph of a 12-layer *n*-eicosanoic acid LB film deposited onto single-crystal silicon [53.98]. Lines of individual molecules are evident at the magnification shown. Figure 53.15 contrasts the molecular organisation expected in an LB multilayer with that in electrostatically deposited layer-by-layer films. For the latter case, the polyelectrolyte chains within each layer will become entangled, and may even penetrate into the layers above and below, leading to a less-ordered film than that produced by LB deposition.

53.4.3 Patterning Technologies

The problem of connecting together the individual processing elements in any future molecular computer is challenging. Each one of the 55 million transistors in the

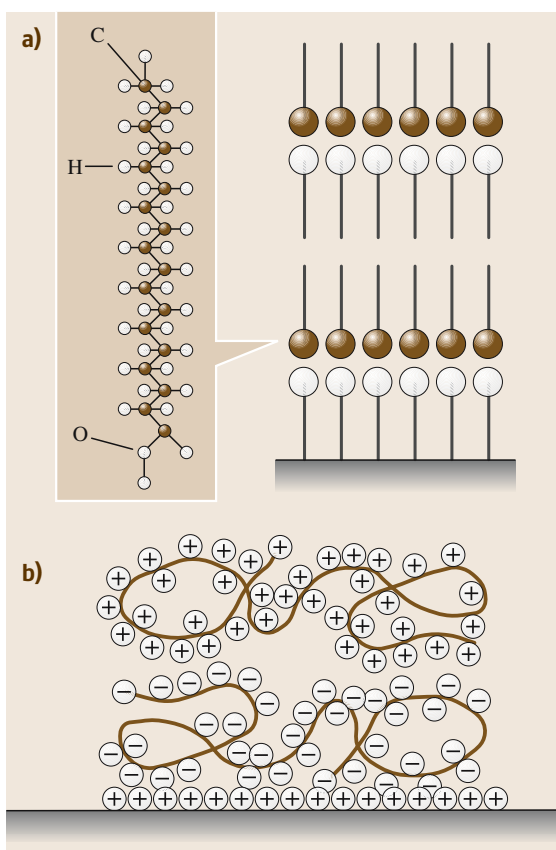


Fig. 53.15a,b Molecular organisation in (a) a Langmuir–Blodgett multilayer assembly and (b) a polyelectrolyte multilayer deposited by the electrostatic method

Pentium®4 chip is addressable and connected to a power supply. Organic molecules can be difficult to arrange on a surface or in a three-dimensional array such that each molecule is addressable. Planar inorganic materials are normally patterned using photolithography. Here, a surface is first covered with a light-sensitive photoresist, which is exposed to ultraviolet light through a contact mask. Either the exposed photoresist (positive resist) or the unexposed regions (negative resist) can then be developed to leave a positive or negative image of the mask on the surface. This approach is routinely used in the fabrication of devices based on inorganic semiconductors. However, difficulties can be encountered when used with organic films, as the photoresists themselves are based on organic compounds.

Brittain et al. [53.99] describe a series of *soft* lithographic methods that may be better suited to the patterning of organic layers. Pouring a liquid polymer,

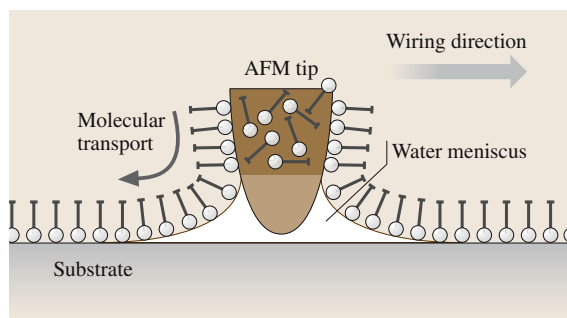


Fig. 53.16 Dip-pen patterning showing the transfer of an *ink* onto *paper* using the tip of an atomic force microscope. (After Piner et al. [53.100])

such as polydimethylsiloxane (PDMS) onto a *master* made from silicon forms a pattern-transfer element. The polymer is allowed to cure to form an elastomer, which can then be removed from the master. This replica can subsequently be used as a stamp to transfer chemical ink, such as a solution of an alkanethiol, to a surface.

Scanning microscopy methods offer a powerful means of manipulating molecules. Careful control of an atomic force microscope (AFM) tip can allow patterns to be drawn in an organic film [53.101, 102]. Such techniques can also be used to reposition molecules, such as the fullerene C_{60} , on surfaces and to break up an individual molecule [53.103]. A further approach that has recently been developed is called dip-pen nanolithography (DPN) [53.100], Fig. 53.16. This is able to deliver organic molecules in a positive printing mode. An AFM tip is used to *write* alkanethiols on a gold thin film in a manner analogous to that of a fountain pen. Molecules flow from the AFM tip to a solid substrate (*paper*) via capillary transport, making DPN a potentially useful tool for assembling nanoscale devices. Recent developments of DPN have included an overwriting capability that allows one nanostructure to be generated and the areas surrounding that nanostructure to be filled with a second type of ink [53.104]. Perhaps the greatest limitation in using scanning probe methodologies for ultrahigh-resolution nanolithography over large areas derives from the serial nature of most techniques. However, an eight-pen nanoplotter capable of doing parallel lithography has been reported [53.105]. The DPN method has also been used to deposit magnetic nanostructures [53.106] and arrays of protein molecules [53.107].

The need to combine large-area coatings with device patterning has resulted in the development of direct-write fabrication methods, such as ink-jet printing [53.108–110]. Although ink-jet printhead droplet

ejection can be achieved with thermal (bubble-jet) and piezoelectric modes of operation, the majority of published literature on ink-jet printing as a tool for manufacturing organic devices has been the result of using piezoelectric-actuated printers. Piezoelectric printhead technology is favoured primarily because it applies no thermal load to the organic *inks* and is compatible with the printing of digital images. The combination of solution-processable emissive polymers with ink-jet printing offers some promise in the development of low-cost high-resolution displays [53.111]. The technique has also been applied to the manufacture of all-polymer transistor circuits [53.112, 113].

53.4.4 Molecular Device Architectures

The *bottom-up* approach to molecular electronics offers many intriguing prospects for manipulating materials on the nanometre scale, thereby providing opportunities to build up novel architectures with predetermined and unique physical and/or chemical properties. Two relatively simple examples of organic superlattice structures are the incorporation of an electric polarisation into a multilayer array to form thin films exhibiting pyroelectric behaviour [53.114] and the use of non-centrosymmetric layers for second-order nonlinear optical response, e.g. second-harmonic generation [53.115]. In both instances, the multilayer thin films can be built up using the **LB** approach.

To realise functional nanoelectronic circuits, a number of workers have investigated the electrical characteristics of structures in which organic molecules are sandwiched between two metallic electrodes. Of particular interest is the possibility of observing molecular rectification using monolayer or multilayer films. This follows the prediction [53.116] that an asymmetric organic molecule containing a donor and an acceptor group separated by a short σ -bonded bridge, allowing quantum-mechanical tunnelling, should exhibit diode characteristics. There have been many attempts to demonstrate this effect in the laboratory, particularly in organic thin films [53.117–119]. Asymmetric current-versus-voltage behaviour has certainly been recorded for many metal/insulator/metal structures, although these results are often open to several interpretations as a result of the asymmetry of the electrode configuration.

In other cases *switching* behaviour has been reported [53.3–8, 12, 13, 119]. As with the work on molecular rectification, the origin of the switching is not always clear. For example, is this a property of the organic molecules, the metallic electrode or of the

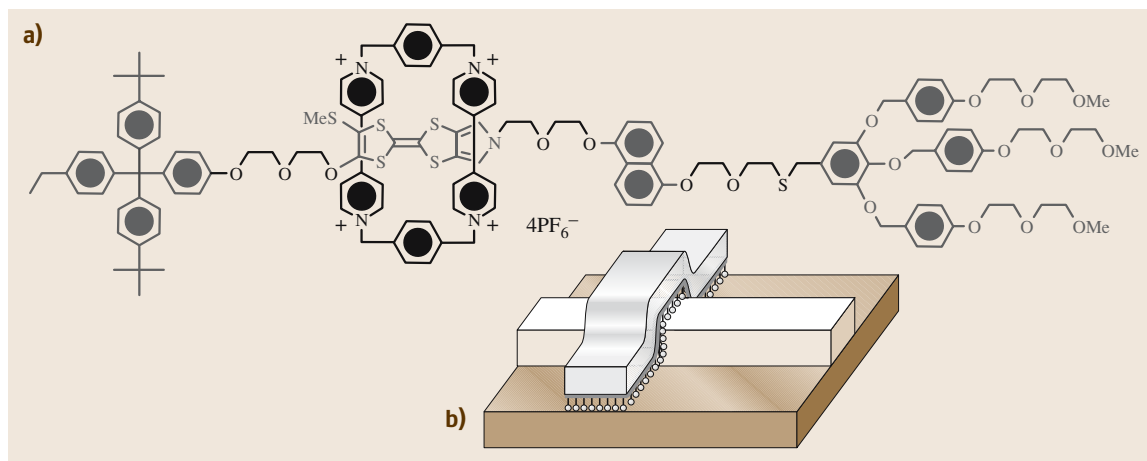


Fig. 53.17 (a) Bistable rotaxane. (b) Schematic diagram of cross-wire structure for switching studies. (After Chen et al. [53.5])

presence of any interfacial (e.g. oxide) layer (or a combination of these)? Bistable rotaxane molecules have been used as the basis of some of these studies, an example is shown in Fig. 53.17a. This molecule is amphiphilic and the ring component can move between the polar and nonpolar regions of the main part of the molecule. The molecules can be assembled on an electrode using the LB approach and a top electrode then deposited to form the crossbar structure shown in Fig. 53.17b [53.5–7]. Alternatively, the molecules can be solution-cast between Pt source and drain electrodes (with gaps of 1–2 nm) in transistor structures [53.9]. In some of the experiments,

the absence of switching using *control* compounds suggests that the rotaxane molecule itself is responsible for the bistability [53.7]. Such results augur well for the development of molecular-scale logic circuitry.

Nanoscale organic devices can also exploit charge storage on nanoparticles or at interfaces. One important metal oxide semiconductor (MOS) device is the flash memory [53.120]. This is similar in structure to a MOS field-effect transistor (MOSFET), except that it has two gate electrodes, one on top of the other. The top electrode forms the control gate, below which a *floating gate* is capacitively coupled to the

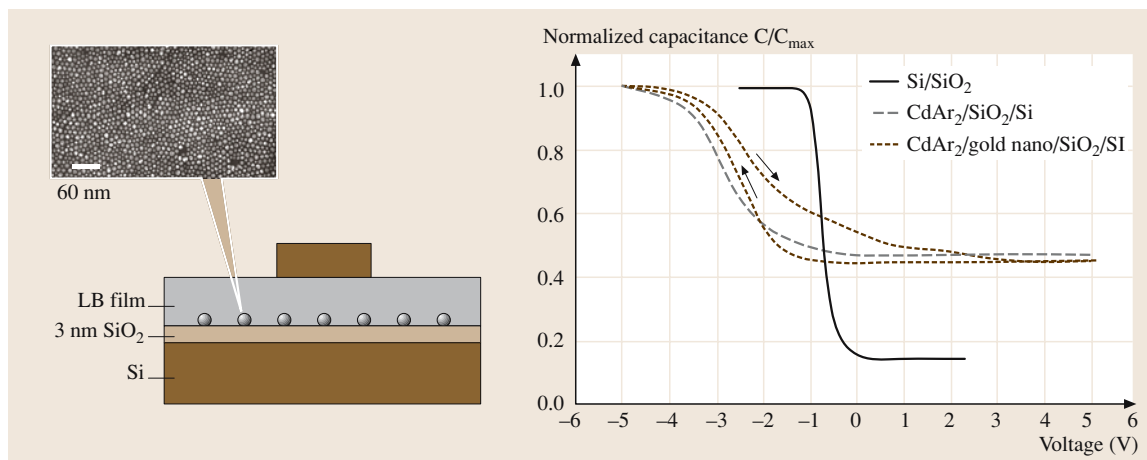


Fig. 53.18 Left: metal–insulator–semiconductor (MIS) structure incorporating gold nanoparticles; a transmission electron micrograph of the nanoparticles is shown. Right: normalised capacitance versus voltage characteristics for different MIS devices. (After Paul et al. [53.10], © (2003) ACS)

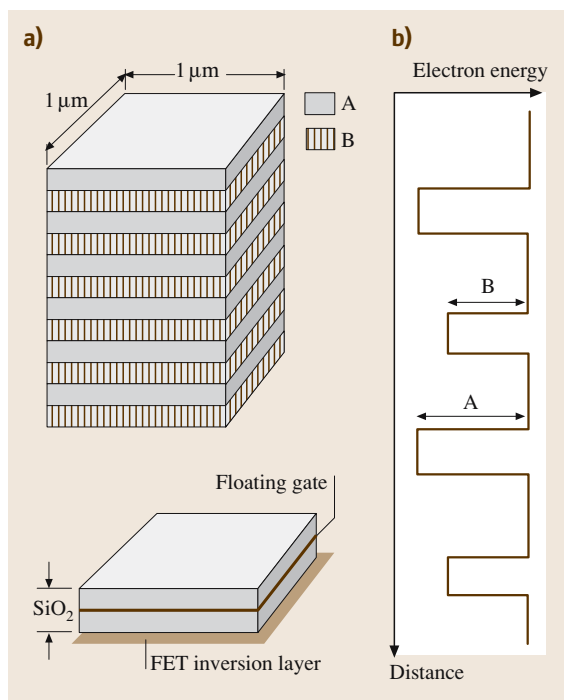


Fig. 53.19 (a) Molecular memory holding N bits (top) compared to a conventional silicon memory holding 1 bit (bottom). (b) Electron energy versus perpendicular distance for molecular memory with no applied electric field. (After Burrows et al. [53.123], with permission, © Elsevier [123])

control gate and the underlying silicon. The memory cell operation involves putting charge on the floating gate or removing it, corresponding to two logic levels. Nanoflash devices utilise single or multiple nanoparticles as the charge-storage elements. These are usually embedded in the gate oxide of a field-effect transistor and located in close proximity (2–3 nm) to the transistor channel [53.121]. Figure 53.18 shows the characteristics of a memory device incorporating metallic nanoparticles deposited by the LB technique [53.10]. The device structure is shown schematically in the figure. The gold nanoparticles (Q-Au) were of nominal

diameter 10 nm and passivated with an organic capping layer. Cadmium arachidate (CdAr_2) LB layers were used to provide an insulating gate layer. Figure 53.18 also shows the normalised capacitance versus voltage ($C-V$) data, measured at 1 MHz and at a voltage scan rate of 40 mV/s for three different device structures: Al/SiO₂/p-Si; Al/20 LB layers CdAr₂/SiO₂/p-Si; and Al/20 LB layers CdAr₂/one LB layer Q-Au/p-Si. The $C-V$ curve for the reference Al/SiO₂/Si sample reveals the usual accumulation/depletion/inversion characteristics associated with metal–insulator–semiconductor (MIS) structures, with a flat-band voltage of approximately -1 V. Negligible hysteresis was evident on reversing the voltage scan. The data for the Si/SiO₂/CdAr₂ structure also show clear accumulation, depletion and inversion regions, again with no hysteresis on reversing the direction of the voltage scan. The most significant difference in the structures with and without the Q-Au nanoparticles is the relatively large hysteresis in the MIS structure containing the Q-Au layer. This was thought to be indicative of charge storage in the gold nanoparticles [53.10]. In a somewhat different approach, the same group has used a self-assembly technique to chemically attach gold nanoparticles to a SiO₂ surface [53.122]. When incorporated into a transistor structure, the resulting device was shown to behave as a nonvolatile electrically erasable programmable read-only memory.

Molecular-scale electronics may also offer increased device densities by fabricating three-dimensional architectures. In 1989, the principle of a three-dimensional memory based on LB films was described [53.123]. The device requires a molecule with a central conjugated region of high electron affinity (for an n-type material, the electron affinity is the energy difference between the bottom of the conduction band and the vacuum level) surrounded by aliphatic substituents of low electron affinity. A multilayer structure, as shown in Fig. 53.19, could be used to store one N -bit word, the presence or absence of charge on the n -th layer representing a 0 or 1 of the n -th bit. The LB film could be assembled on the gate of an FET and, on application of an electric field, transport of bits across the layers may be detected as induced charge on the gate.

53.5 DNA Electronics

The study of the electronic behaviour of organic compounds has led some scientists to work on the electrical properties of biological materials. Deoxyribonucleic

acid (DNA) is arguably the most significant molecule in nature. It may also be an important material for molecular electronics applications. Reports into the

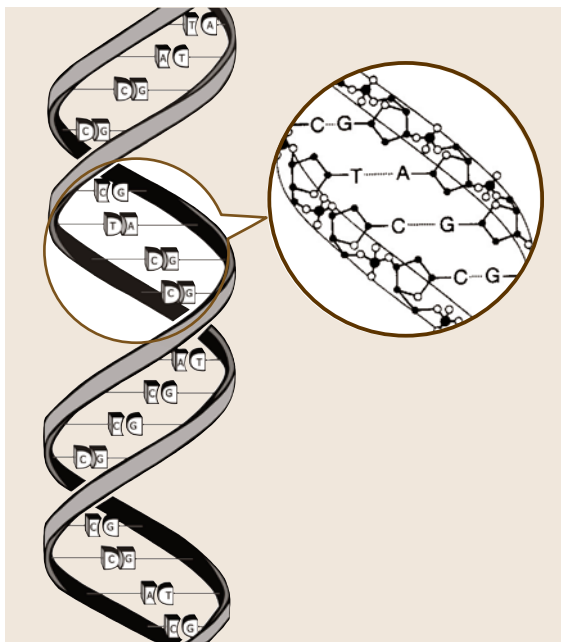


Fig. 53.20 DNA double helix showing the position of the four bases: guanine (G), cytosine (C), adenine (A) and thymine (T)

electronic properties of DNA have already generated controversy in the literature [53.16, 124–128]. According to some, DNA is a molecular wire of very small resistance. Others, however, find that DNA behaves as an insulator. These seemingly contradictory findings can probably be explained by the different DNA sequences and experimental conditions used to monitor the conductivity [53.16].

The DNA strand in the double-helix arrangement consists of a long polymer backbone consisting of repeating sugar molecules and phosphate groups, Fig. 53.20. Each sugar group is attached to one of four bases, guanine (G), cytosine (C), adenine (A) and thymine (T). The chemical bonding is such that an A base only ever pairs with a T base, while a G always pairs with a C. Some of the electron orbitals belonging to the bases overlap quite well with each other along the long axis of the DNA. This provides a path for elec-

tron transfer along the molecule, in a similar fashion to the one-dimensional conduction seen in conjugated polymers, Sect. 53.1.2.

Theoretical and experimental work now suggests that a hole (i. e. a positive charge) is more stable on a G–C base pair than on an A–T base pair [53.16]. The energy difference between these two pairs is substantially larger than the thermal energy of the charge carrier. Under these conditions, a hole will localise on a particular G–C base pair. Because the A–T base pairs have a higher energy, they act as a barrier to hole transfer. However, if the distance between two G–C base pairs is small enough, the hole can tunnel quantum-mechanically from one pair to the next. In this way, charge carriers are able to shuttle along a single DNA molecule over a distance of a few nanometres.

DNA chips exploit the fact that short strands of DNA will bind to other segments of DNA that have complementary sequences, and can therefore be used to probe whether certain genetic codes are present in a given specimen of DNA. Microfabricated chips with many parallel DNA probes are becoming widespread in analytical and medical applications. Currently, the chips are read out optically, but further miniaturisation might require new read-out schemes, possibly involving the electron-transfer properties of DNA.

Computations by chemical or biological reactions overcome the problem of parallelism and interconnections in a classical system. If a string of DNA can be put together in the right sequence, it can be used to solve combinatorial problems. The calculations are performed in test tubes filled with strands of DNA. Gene sequencing is used to obtain the result. For example, *Adleman* [53.129] calculated the *travelling-salesman problem* to demonstrate the capabilities of DNA computing. DNA computing on parallel problems potentially provides 10^{14} millions of instructions per second (MIPS) and uses less energy and space than conventional supercomputers. While CMOS supercomputers operate 10^9 operations per Joule, a DNA computer could perform about 10^{19} operations per Joule. Data could potentially be stored on DNA in a density of approximately 1 bit per nm^3 while existing storage media such as DRAMs require 10^{12} nm^3 to store 1 bit.

53.6 Conclusions

Organic compounds possess a wide range of fascinating physical and chemical properties that make them attractive candidates for exploitation in electronic and

optoelectronic devices. It is not, however, anticipated that these materials will displace silicon in the foreseeable future as the dominant material for fast signal

processing. It is much more likely that organic materials will find use in other niche areas of electronics, where silicon and other inorganic semiconductors cannot compete. Examples already exist, such as liquid-crystal displays and certain chemical sensors. Organic light-emitting structures are likely to make a major impact in the marketplace over the next ten years.

Over the first decades of the 21st century, classical CMOS technology will come up against a number of technological barriers. The bottom-up approach to molecular electronics provides an alternative and attrac-

tive way forward and, as such, it is currently an area of exciting interdisciplinary activity. However, the challenges in fabricating molecular switches and connecting them together are formidable. Living systems use a different approach; these assemble themselves naturally from molecules and are extremely energetically efficient when compared with man-made computational devices. More radical approaches to materials fabrication and device design, exploiting self-organisation, may be needed to realise fully the potential offered by molecular-scale electronics.

References

- 53.1 M. C. Petty, M. R. Bryce, D. Bloor (eds): *An Introduction to Molecular Electronics* (Edward Arnold, London 1995)
- 53.2 T. H. Richardson (ed): *Functional Organic and Polymeric Materials* (Wiley, Chichester 2000)
- 53.3 J. M. Tour: *Molecular Electronics* (World Scientific, New Jersey 2003) Chap. 2
- 53.4 P. E. Kornilovitch, A. M. Bratkovsky, R. S. Williams: *Phys. Rev. B* **66**, 245413 (2002)
- 53.5 Y. Chen, G.-Y. Jung, D. A. A. Ohlberg, X. Li, D. R. Stewart, J. O. Jeppesen, K. A. Nielsen, J. F. Stoddart, R. S. Williams: *Nanotechnology* **14**, 462 (2003)
- 53.6 Y. Chen, D. A. A. Ohlberg, X. Li, D. R. Stewart, R. S. Williams, J. O. Jeppesen, K. A. Nielsen, J. F. Stoddart, D. L. Olynick, E. Anderson: *Appl. Phys. Lett.* **82**, 1610 (2003)
- 53.7 M. R. Diehl, D. W. Steuerman, H.-R. Tseng, S. A. Virginon, A. Star, P. C. Celestre, J. F. Stoddart, J. R. Heath: *ChemPhysChem* **4**, 1335 (2003)
- 53.8 L. Ma, S. Pyo, J. Ouyang, Q. Xu, Y. Yang: *Appl. Phys. Lett.* **82**, 1419 (2003)
- 53.9 H. Yu, Y. Luo, K. Beverly, J. F. Stoddart, H.-R. Tseng, J. R. Heath: *Angew. Chem.* **42**, 5706 (2003)
- 53.10 S. Paul, C. Pearson, A. Molloy, M. A. Cousins, M. Green, S. Kolliopoulou, P. Dimitrakis, P. Normand, D. Tsoukalas, M. C. Petty: *Nano. Lett.* **3**, 533 (2003)
- 53.11 J. M. Tour, L. Cheng, D. P. Nackashi, Y. Yao, A. K. Flatt, S. K. St. Angelo, T. E. Mallouk, P. D. Franzén: *J. Am. Chem. Soc.* **125**, 13279 (2003)
- 53.12 L. D. Bozano, B. W. Kean, V. R. Deline, J. R. Salem, J. C. Scott: *Appl. Phys. Lett.* **84**, 607 (2004)
- 53.13 J. Ouyang, C.-W. Chu, C. R. Szmanda, L. Ma, Y. Yang: *Nature Mater.* **3**, 918 (2004)
- 53.14 R. M. Metzger: *J. Solid State Chem.* **168**, 696 (2002)
- 53.15 J. Chen, W. Wang, M. A. Reed, A. M. Rawlett, D. W. Price, J. M. Tour: *Appl. Phys. Lett.* **77**, 1224 (2000)
- 53.16 C. Dekker, M. A. Ratner: *Phys. World* **14**, 29 (2001)
- 53.17 K. E. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley, New York 1992)
- 53.18 C. Nicolini (ed.): *Molecular Manufacturing* (Plenum, New York 1996)
- 53.19 S. Roth: *One-Dimensional Metals* (VCH, Weinheim 1995)
- 53.20 C. D. Dimitrakopoulos, D. J. Masearo: *IBM J. Res. Devel.* **45**, 11 (2001)
- 53.21 M. R. Bryce, M. C. Petty: *Nature* **374**, 771 (1995)
- 53.22 W. J. Feast, J. Tsibouklis, K. L. Pouwer, L. Groenen-daal, E. W. Meijer: *Polymer* **37**, 5017 (1996)
- 53.23 J. R. Ferraro, J. M. Williams: *Introduction to Synthetic Electrical Conductors* (Academic, Orlando 1987)
- 53.24 J. D. Wright: *Molecular Crystals* (Cambridge Univ. Press, Cambridge 1995)
- 53.25 H. W. Kroto, D. R. M. Walton (eds): *The Fullerenes* (Cambridge Univ. Press, Cambridge 1993)
- 53.26 G. Timp (ed.): *Nanotechnology* (Springer, Berlin Heidelberg New York 1998)
- 53.27 R. Saito, G. Dresselhaus, M. S. Dresselhaus: *Physical Properties of Carbon Nanotubes* (Imperial College Press, London 1998)
- 53.28 L. Duclaux: *Carbon* **40**, 1751 (2002)
- 53.29 P. May: *Phys. World* **8**, 52 (1995)
- 53.30 Plastic Logic, Cambridge, UK (2005): <http://www.plasticlogic.com/>
- 53.31 S. R. Forrest: *Nature* **428**, 911 (2004)
- 53.32 E. H. Rhoderick: *Metal-Semiconductor Contacts* (Clarendon, Oxford 1978)
- 53.33 Y. L. Hua, M. C. Petty, G. G. Roberts, M. M. Ahmad, M. Hanack, M. Rein: *Thin Solid Films* **149**, 161 (1987)
- 53.34 G. Horowitz: *Adv. Mater.* **2**, 286 (1990)
- 53.35 H. E. Katz: *J. Mater. Chem.* **7**, 369 (1997)
- 53.36 S. Scheinert, G. Paasch: *Phys. Status Solidi (a)* **201**, 1263–1301 (2004)
- 53.37 A. R. Brown, A. Pomp, C. M. Hart, D. M. de Leeuw: *Science* **270**, 972 (1995)

- 53.38 J. G. Laquindanum, H. E. Katz, A. J. Lovinger, A. Dodabalapur: *Chem. Mater.* **8**, 2542 (1996)
- 53.39 Z. Bao, A. J. Lovinger, A. Dodabalapur: *Adv. Mater.* **9**, 42 (1997)
- 53.40 C. D. Dimitrakopoulos, B. K. Furman, T. Graham, S. Hegde, S. Purushothaman: *Synth. Met.* **92**, 47 (1998)
- 53.41 C. D. Dimitrakopoulos, S. Purushothaman, J. Kymissis, A. Callegari, J. M. Shaw: *Science* **283**, 822 (1999)
- 53.42 H. Sirringhaus, T. Kawase, R. H. Friend, T. Shimoda, M. Inbasekaran, W. Wu, E. P. Woo: *Science* **290**, 2123 (2000)
- 53.43 H. Klauk, M. Halik, U. Zschieschang, F. Eder, G. Schmid, C. Dehm: *Appl. Phys. Lett.* **82**, 4175 (2003)
- 53.44 J. Lee, K. Kim, J. H. Kim, S. Im, D.-Y. Jung: *Appl. Phys. Lett.* **82**, 4169 (2003)
- 53.45 P. F. Baude, D. A. Ender, M. A. Haase, T. W. Kelley, D. V. Muyres, S. D. Theiss: *Appl. Phys. Lett.* **82**, 3964 (2003)
- 53.46 M. S. Fuhrer, B. M. Kim, T. Durkop, T. Brintlinger: *Nano. Lett.* **2**, 755 (2002)
- 53.47 A. Javey, Q. Wang, A. Ural, Y. M. Li, H. J. Dai: *Nano. Lett.* **2**, 929 (2002)
- 53.48 F. Léonard, J. Tersoff: *Phys. Rev. Lett.* **88**, 258302-1 (2002)
- 53.49 K. Tsukagoshi, N. Yoneya, S. Uryu, Y. Aoyagi, A. Kanda, Y. Ootuka, B. W. Alphenaar: *Physica B* **323**, 107 (2002)
- 53.50 J. H. Burroughes, D. D. C. Bradley, A. R. Brown, R. N. Marks, K. Mackay, R. H. Friend, P. L. Burns, A. B. Holmes: *Nature* **347**, 359 (1990)
- 53.51 S. Miyata, H. S. Nalwa (eds): *Organic Electroluminescent Materials and Devices* (Gordon Breach, Amsterdam 1997)
- 53.52 A. J. Hudson, M. S. Weaver: *Organic electroluminescence*. In: *Functional Organic and Polymeric Materials*, ed. by T. H. Richardson (Wiley, Chichester 2000) p. 365
- 53.53 R. Farchioni, G. Grosso (eds): *Organic Electronic Materials* (Springer, Berlin, Heidelberg 2001)
- 53.54 Cambridge Display Technology (data provided February 2005): <http://www.cdtltd.co.uk/>
- 53.55 J. K. Borchardt: *Mater. Today* **7**, 42 (Sept. 2004)
- 53.56 J. H. Ahn, C. Wang, C. Pearson, M. R. Bryce, M. C. Petty: *Appl. Phys. Lett.* **85**, 1283 (2004)
- 53.57 L. S. Hung, C. W. Tang, M. G. Mason: *Appl. Phys. Lett.* **70**, 152 (1997)
- 53.58 G.-Y. Jung, C. Pearson, L. E. Horsburgh, I. D. W. Samuel, A. P. Monkman, M. C. Petty: *J. Phys. D: Appl. Phys.* **33**, 1029 (2000)
- 53.59 Y. Cao, I. D. Parker, G. Yu, C. Zhang, A. J. Heeger: *Nature* **397**, 414 (1999)
- 53.60 P. Cea, Y. Hua, C. Pearson, C. Wang, M. R. Bryce, M. C. López, M. C. Petty: *Mater. Sci. Eng. C* **22**, 87 (2002)
- 53.61 P. Cea, Y. Hua, C. Pearson, C. Wang, M. R. Bryce, F. M. Royo, M. C. Petty: *Thin Solid Films* **408**, 275 (2002)
- 53.62 M. Ikai, S. Tokito, Y. Sakamoto, T. Suzuki, Y. Taga: *Appl. Phys. Lett.* **79**, 156-158 (2001)
- 53.63 J. Kido, M. Kimura, K. Nagai: *Science* **267**, 1332 (1995)
- 53.64 J. Kido, H. Shionoya, K. Nagai: *Appl. Phys. Lett.* **67**, 2281 (1995)
- 53.65 J. Thompson, R. I. R. Blyth, M. Mazzeo, M. Anni, G. Gigli, R. Cingolani: *Appl. Phys. Lett.* **79**, 560 (2001)
- 53.66 A. R. Duggal, J. J. Shiang, C. M. Heller, D. F. Forest: *Appl. Phys. Lett.* **80**, 3470 (2002)
- 53.67 N. S. Saricifti: *Mater. Today* **7**, 36 (Sept. 2004)
- 53.68 B. O'Regan, M. Grätzel: *Nature* **353**, 737 (1991)
- 53.69 A. Hinsch, J. Kroon, R. Kern, I. Uhlenndorf, R. Sastrowan, A. Meyer: *Proc. 17th Eur. Photov. Solar Energy Conf.* **51**, WIP-Munich and ETA-Florence (2001)
- 53.70 J. Janata: *Principles of Chemical Sensors* (Plenum, New York 1989)
- 53.71 J. W. Gardner: *Microsensors* (Wiley, Chichester 1994)
- 53.72 J. W. Gardner, V. K. Varadan, O. O. Awadelkarim: *Microsensors, MEMS and Smart Devices: Technology, Applications and Devices* (Wiley, Chichester 2001)
- 53.73 P. T. Moseley, A. J. Crocker: *Sensor Materials* (IOP, Bristol 1996)
- 53.74 A. S. Snow, W. R. Barger: *Phthalocyanine films in chemical sensors*. In: *Phthalocyanines: Properties and Applications*, ed. by C. C. Leznoff A. B. P. Lever (VCH, Weinheim 1989) p. 342
- 53.75 M. C. Petty, R. Casalini: *Eng. Science and Education Journal* **10**, 99 (2001)
- 53.76 S. Baker, G. G. Roberts, M. C. Petty: *IEE Proc.* **130**, 260 (1983)Pt. 1
- 53.77 S. D. Senturia, C. M. Sechen, J. A. Wishneusky: *Appl. Phys. Lett.* **30**, 106 (1977)
- 53.78 P. S. Barker, C. Di Bartolomeo, A. P. Monkman, M. C. Petty, R. Pride: *Sensors Actuators B* **25**, 451 (1995)
- 53.79 P. S. Barker, M. C. Petty, A. P. Monkman, J. McMurdo, M. J. Cook, R. Pride: *Thin Solid Films* **284-285**, 94 (1996)
- 53.80 J. W. Gardner, P. N. Bartlett: *Electronic Noses: Principles and Applications* (Oxford Univ. Press, Oxford 1999)
- 53.81 Semiconductor Industry Association Roadmap (2003): <http://public.itrs.net/>
- 53.82 J. Greer, A. Korkin, J. Lanbanowski (eds): *Nano and Giga Challenges in Microelectronics* (Elsevier, Amsterdam 2003)
- 53.83 R. Dettmer: *IEE Rev.* **49**, 30 (July 2003)
- 53.84 M. C. Petty: *Organic thin-film deposition techniques*. In: *Functional Organic and Polymeric Materials*, ed. by T. H. Richardson (Wiley, Chichester 2000) p. 7
- 53.85 A. Ulman: *Ultrathin Organic Films* (Academic, San Diego 1991)
- 53.86 M. C. Petty: *Langmuir-Blodgett Films* (Cambridge Univ. Press, Cambridge 1996)

- 53.87 D. R. Talham: Chem. Rev. **104**, 5479 (2004)
- 53.88 A. J. Moore, L. M. Goldenberg, M. R. Bryce, M. C. Petty, A. P. Monkman, C. Marenco, J. Yarwood, M. J. Joyce, S. N. Port: Adv. Mater. **10**, 395 (1998)
- 53.89 L. Netzer, J. Sagiv: J. Am. Chem. Soc. **105**, 674 (1983)
- 53.90 L. Netzer, R. Iscovici, J. Sagiv: Thin Solid Films **99**, 235 (1983)
- 53.91 G. Decher, J. D. Hong, J. Schmitt: Thin Solid Films **210/211**, 831 (1992)
- 53.92 G. Decher, Y. Lvov, J. Schmitt: Thin Solid Films **244**, 772 (1994)
- 53.93 G. Decher, J. B. Schlenoff (eds): *Multilayer Thin Films* (Wiley-VCH, Weinheim 2003)
- 53.94 C. Pearson, J. Nagel, M. C. Petty: J. Phys. D: Appl. Phys. **34**, 285 (2001)
- 53.95 J. H. Cheung, W. B. Stockton, M. F. Rubner: Macromolecules **30**, 2712 (1997)
- 53.96 Y. Lvov, M. Onda, K. Ariga, T. Kunitake: J. Biomater. Sci. Polym. Edn. **9**, 345 (1998)
- 53.97 W. B. Stockton, M. F. Rubner: Macromolecules **30**, 2717 (1997)
- 53.98 S. A. Evanson, J. P. S. Badyal, C. Pearson, M. C. Petty: J. Phys. Chem. **100**, 11672 (1996)
- 53.99 S. Brittain, K. Paul, X.-M. Zhao, G. Whitesides: Phys. World **11**, 31 (1998)
- 53.100 R. D. Piner, J. Zhu, F. Xu, S. Hong, C. A. Mirkin: Science **283**, 661 (1999)
- 53.101 L. F. Chi, L. M. Eng, K. Graf, H. Fuchs: Langmuir **8**, 2255 (1992)
- 53.102 A. Krämer, R. R. Fuierer, C. B. Gorman: Chem. Rev. **103**, 4367 (2003)
- 53.103 J. Gimzewski: Phys. World **11**, 25 (1998)
- 53.104 S. Hong, J. Zhu, C. A. Mirkin: Science **286**, 523 (1999)
- 53.105 S. Hong, C. A. Mirkin: Science **288**, 1808 (2000)
- 53.106 X. G. Liu, L. Fu, S. H. Hong, V. P. Dravid, C. A. Mirkin: Adv. Mater. **14**, 231 (2002)
- 53.107 K. B. Lee, S. J. Park, C. A. Mirkin, J. C. Smith, M. Mrksich: Science **295**, 1702 (2002)
- 53.108 S. P. Speakman, G. G. Rozenberg, K. J. Clay, W. I. Milne, A. Ille, I. A. Gardner, E. Bresler, J. H. G. Steinke: Organic Electron. **2**, 65 (2001)
- 53.109 G. PerHin, B. T. Khuri-Yakub: Rev. Sci. Instrum. **73**, 2193 (2002)
- 53.110 H. Sirringhaus, T. Shimoda (eds): MRS Bull. **28**, 802 (2003)
- 53.111 I. D. Rees, K. L. Robinson, A. B. Holmes, C. R. Towns, R. O'Dell: MRS Bull. **27**, 451 (2002)
- 53.112 H. Sirringhaus, T. Kawase, R. H. Friend: MRS Bull. **26**, 539 (2001)
- 53.113 K. E. Paul, W. S. Wong, S. E. Ready, R. A. Street: Appl. Phys. Lett. **83**, 2070 (2003)
- 53.114 C. A. Jones, M. C. Petty, G. G. Roberts: IEEE Trans. Ultrasonics Ferroelec. Freq. Control **35**, 736 (1988)
- 53.115 P. J. Skabara: Organic second-order non-linear optical materials. In: *Functional Organic and Polymeric Materials*, ed. by T. H. Richardson (Wiley, Chichester 2000) p. 295
- 53.116 A. Aviram, M. A. Ratner: Chem. Phys. Lett. **29**, 277 (1974)
- 53.117 A. S. Martin, J. R. Sambles, G. J. Ashwell: Phys. Rev. Lett. **70**, 218 (1993)
- 53.118 G. J. Ashwell, D. S. Gandolfo: J. Mater. Chem. **11**, 246 (2001)
- 53.119 J. R. Heath, P. J. Kuekes, G. S. Snider, R. S. Williams: Science **280**, 1716 (1998)
- 53.120 B. G. Streetman, S. Banjeree: *Solid State Electronic Devices* (Prentice Hall, New Jersey 2000)
- 53.121 R. Compañó (ed): *Technology Roadmap for Nanoelectronics* (Office for Official Publications of the European Communities, Luxembourg 2001)
- 53.122 S. Kolliopoulou, P. Dimitrakakis, P. Normand, H. -L. Zhang, N. Cant, S. D. Evans, S. Paul, C. Pearson, A. Molloy, M. C. Petty: J. Appl. Phys. **94**, 5234 (2003)
- 53.123 P. E. Burrows, K. J. Donovan, E. G. Wilson: Thin Solid Films **179**, 129 (1989)
- 53.124 H.-W. Fink, C. Schöneberger: Nature **398**, 407 (1999)
- 53.125 R. S. Phadke: Appl. Biochem. Biotechnol. **96**, 269 (2001)
- 53.126 A. Rakitin, P. Aich, C. Papadopoulos, Y. Kobzar, A. S. Vedeneev, J. S. Lee, J. M. Xu: Phys. Rev. Lett. **86**, 3670 (2001)
- 53.127 C. N. R. Rao, A. K. Cheetham: J. Mater. Chem. **11**, 2887 (2001)
- 53.128 H. Tabata, L. T. Cai, J. H. Gu, S. Tanaka, Y. Otsuka, Y. Sacho, M. Taniguchi, T. Kawai: Synth. Met. **133-134**, 469 (2003)
- 53.129 L. Adleman: Science **266**, 1021 (1994)

54. Organic Materials for Chemical Sensing

Organic materials for chemical sensing are broadly classified into three categories: (i) macrocyclic compounds, (ii) conducting polymers and (iii) cavitand molecules. A short review of current progress in semiconductor oxide sensing materials is given, pointing out their strengths and limitations. Principal wet techniques for depositing organic thin films are described and electrical, optical and structural properties of all three types of organic materials are analysed in relation to their importance in chemical sensing. Examples of recent advances in chemical sensing of different analytes and pollutants are presented.

54.1	Analyte Requirements	1242	54.3.1	Preparation of Sensing Membranes	1245
54.2	Brief Review of Inorganic Materials	1243	54.3.2	Thin-Film Properties	1248
54.3	Macrocyclic Compounds for Sensing	1245	54.4	Sensing with Phthalocyanine and Porphyrin	1250
			54.4.1	Amperometric Sensor	1250
			54.4.2	Optical Sensors	1251
			54.4.3	Detection of Volatile Organic Vapour Compounds	1254
			54.5	Polymeric Materials	1255
			54.5.1	Conducting Polymers	1255
			54.5.2	Ion Sensing	1257
			54.5.3	Examples of Other Polymeric Sensors	1257
			54.6	Cavitand Molecules	1259
			54.7	Concluding Remarks	1261
			References	1262

The development of a new generation of sensors involves the study of fine-tuned sensor-active materials and transducers [54.1, 2]. Metal oxides, polymer/polymer composites, and dyes are being regarded as key sensing materials. This chapter is organised into eight sections. The next section presents a summary of the needs for chemical sensing of different pollutants. The second section provides a brief description of different semiconducting oxide materials and their applications in sensor technology. Organised ultrathin organic films of molecular and polymeric semiconductors are currently the focus of considerable research with an emphasis on a molecular understanding of the sensing mechanisms. The low cost of fabrication, large-area processability, and widely diverse and tailorable physical, electronic, and optical properties have stimulated interest in their practical exploitation. Phthalocyanine and the related macrocyclic compounds have been extensively investigated for chemical sensing primarily because the desired functionality of phthalocyanine molecules can be achieved by changing the central metal atom or introducing substituents in their π -electron aromatic ring [54.3]. The third section is devoted to thin-film formulation of these

interesting compounds and their structural, electrical and optical properties relevant to sensing different types of pollutants. Important transduction methods and their applications in the development of practical sensors are presented in the fourth section, with an emphasis on the interaction at the interface of analyte/film of macrocyclic compounds. The ability of conducting polymers to change their physical properties during reaction with various redox agents makes them very useful for the development of gas sensors. The most common are polyaniline (PAni) and polypyrrole (PPY). It has been shown that the resistivity of polypyrrole increases in the presence of a reducing gas such as ammonia, but decreases in the presence of an oxidising gas such as nitrogen dioxide. The gases cause a change in the near-surface charge-carrier density by reacting with surface-adsorbed oxygen ions [54.4]. Sensing properties of polymers and their importance in the development of intelligent sensors are described in Sect. 54.5. A class of supramolecules, called cavitands, that exhibit discotic phases depending on the structure, are shown to have potential for sensing applications [54.5]. Section 54.6 deals with the applications of thin films of these

molecules such as calixresorcinarene for recognition of organic vapours. The chapter concludes by highlighting

the trends of future development of sensor materials and technology in Sect. 54.7.

54.1 Analyte Requirements

The detection of pollutant gases and volatile organic compounds (VOC) whether in gaseous phase or dissolved in water is of great environmental importance due to the extreme hazards posed by their presence in small amounts in the ambient. Nitrogen dioxide (NO₂), ozone (O₃), sulphur dioxide (SO₂), particulate matter (PM), carbon monoxide (CO), and lead (Pb) are six principal pollutants of major concern for maintaining the quality of air in urban areas (Table 54.1). Some of these pollutants, like CO, Pb, NO₂, PM and SO₂, are emitted directly from a variety of sources including industrial production growth and increased traffic congestion. Ozone is generally formed when emissions of NO₂, SO₂, ammonia (NH₃), VOCs, and other gases react in the atmosphere. The presence of ozone in high concentration is also possible in offices where laser printers and photocopiers are heavily used. On the other hand, water is commonly treated with O₃ to destroy microbial contaminants such as *Escherichia coli* and *salmonella* [54.6, 7] and there is a need for accurate calibration of ozone content. Due to its ability in absorbing ultraviolet (UV) radiation from sunlight, stratospheric ozone filters out damaging ultraviolet radiation from the sun. Monitoring of ozone in the stratosphere has thus become a global issue as it forms a protective shield for all forms of life on Earth.

Stringent domestic and international standard for emission control requires the monitoring of threshold

limit value of these pollutants in order to protect primarily against adverse public health hazards. For example VOCs and nitric oxide (NO_x) resulting from combustion processes cause respiratory problems. Benzene itself is regarded as a human carcinogen and vehicle exhaust is primarily composed of these toxic gases. Spillage in fuel stations also contribute to the emission of excessive amounts of aromatic hydrocarbons in the atmosphere. VOCs also bind with ground-level ozone and form a photochemical smog, causing breathing difficulty for asthma sufferers. The secondary requirement of pollution monitoring is to reduce disastrous effects on the preservation of ecosystems, plants and animals. Safeguard is also to be taken against the decreased visibility due to the presence of PM and damage to crops, vegetation, and buildings due to atmospheric pollution. The SO₂ content is known to attack the indoors and outdoors of buildings, and in particular the surface of historical glasses. A wide variety of preventive conservation tools and routines have been sought to protect historical and cultural objects against deterioration and degradation [54.8, 9].

The activities of chemical sensing are not solely limited to environmental control. Hydrogen (H₂) fuel cells are clean, quiet, more efficient generators of electricity than any other known technology. Hydrogen is also an important raw material for the aerospace, chemical and semiconductor industries. As hydrogen is explosive above the lower explosive limit (LEL), H₂ sensors have

Table 54.1 Air quality standards for the UK, USA and World Health Organisation (WHO)

Pollutant	UK		WHO		USA	
	Concentration	Standard measured	Concentration	Standard measured	Concentration	Standard measured
Benzene	5 ppb	Annual mean	Data not available			
1,3-butadiene	1 ppb	Annual mean	Data not available			
Carbon monoxide (CO)	8.6 ppm	8 h mean	8.6 ppm	8 h mean	9 ppm	8 h mean
Lead (Pb)	0.25 µg/m ³	Annual mean	0.5 µg/m ³	Annual mean	1.5 µg/m ³	Quarterly mean
Nitrogen dioxide (NO ₂)	21 ppb	Annual mean	105 ppb	1 h mean	0.05 ppm	Annual mean
Ozone (O ₃)	50 ppb	8 h mean	60 ppb	8 h mean	0.08 ppm	8 h mean
Particles (PM ₁₀)	50 µg/m ³	24 h mean	70 µg/m ³	24 h mean	50 µg/m ³	Annual mean
Sulphur dioxide (SO ₂)	100 ppb	15 min mean	188 ppb	10 min mean	0.14 ppm	24 h mean

now become important safety devices in all these applications. The oxygen sensor which detects the air–fuel mixture of a gasoline engine by measuring the amount of oxygen in the exhaust gas is an essential component of the car engine management system. The development of reliable oxygen sensors for medical application has received considerable research attention in recent years [54.10–12]. Drinking, industrial and swimming-pool water must be disinfected with suitable oxidising agents such as chlorine or chlorine compounds. The dosing of the appropriate oxidising agents must be carefully controlled to suit the application as a higher concentration can result in corrosion effects, impairment of taste or skin irritation. The ability of detection of chlorine in sub-ppm level is therefore essential [54.13]. Ammonia is increasingly used as a refrigerant alternative to ozone-depleting chlorofluorocarbon (CFC). Large quantities of ammonia can also be found in fertiliser factories, resin production plants using urea, explosives munitions plants, semiconductor industries and water utility facilities. The presence of ammonia can cause inhalation problem at a very low concentration (0.6–53 ppm). Airborne ammonia gas dissolves in moisture on the skin, forming corrosive ammonium hydroxide. The exposure of ammonia at 10 000 ppm is mildly irritating to moist skin but the effects become more pronounced with the increasing concentration, producing chemical burns with blistering at 30 000 ppm. Its solubility in water can cause it to cauterise respiratory tracts, resulting in deaths at concentrations of 5000 ppm. Research into the development of reliable ammonia sensors with

a fast response to a wide range of concentrations is intensive [54.14–16].

It is obvious from the brief survey above that growing demand exists for the fabrication of low-cost, low-power, robust and portable chemical sensors for a variety of pollutants, contaminants and analytes, showing improved sensitivity and selectivity for industrial, healthcare and environmental control. Analytical chemical tools provide advanced and reliable methods for detection of different analytes but those methods are in many cases prohibitively expensive, unsuited for large-scale field measurements or mobile monitoring. For instance, there is currently a good market for detection systems for ozone at low concentration (ppb level) to replace more expensive techniques such as UV absorption measurements [54.17]. Existing devices for the detection of ozone are now mostly based on electrochemical measurements [54.18], or on solid-state semiconductor devices [54.19]. Similarly, the detection of organic vapours involves extracting air samples and subsequently analysing them using standard laboratory techniques such as gas chromatography [54.20], mass spectroscopy and Fourier-transform infrared (FTIR) spectroscopy [54.21, 22]. The major disadvantages associated with these techniques are their complexity, the capital cost of the instruments used and the requirement for trained human resources for operation and the interpretation of data. Apart from environmental applications, the development of detection instrumentation of these species is also important for both human health monitoring [54.23, 24] and for odour detection [54.25].

54.2 Brief Review of Inorganic Materials

Solid-state semiconductor gas sensors are usually made of tin oxide (SnO_2) in the form of thick films, porous pellets or thin-film coatings to monitor the presence of various gases [54.26]. These sensors are being used in the intelligent automatic control of a large number of processes, ranging from microwave cooking to the efficient combustion of motor engines. SnO_2 sensors operate at high temperatures of 300–400 °C. Gases are adsorbed onto the surface in the presence of air and the semiconductivity of the metal oxide changes due to the formation of a surface depletion layer. The thickness of this layer can be expressed in terms of the Debye length $L_D = \sqrt{\epsilon_0 \epsilon k_B T / (q^2 n_0)}$, where ϵ_0 is the permittivity of free space, ϵ is the dielectric constant of the semiconducting material, n_0 is the total carrier concentration, q is the electronic charge, k_B is the Boltzmann constant,

and T is the absolute temperature [54.27]. The sensitivity of the sensor is generally defined as the ratio of the change in the conductance after the treatment to the conductance of the sensor in air and can simply written as $\Delta n L_D / n_0$, where Δn is the change in carrier concentration due to the analyte exposure. It is therefore possible to optimise the sensitivity by reducing grain size, lowering the carrier concentration and increasing the Debye length. The maximum sensitivity is assumed to have been achieved when the Debye length is about half the particle size.

Figure 54.1 shows a photograph of a comb-like interdigitated electrode (IDE) system used for most conductivity measurements. The original idea was developed by Taguchi [54.28] and the sample resistance in this configuration is smaller than that obtained from

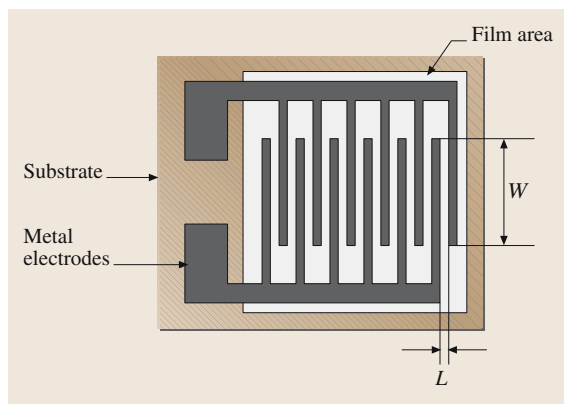


Fig. 54.1 Interdigitated electrode (IDE) system for an SnO₂ sensor

the four-contact geometry. The output signal is normally logarithmic to gas concentration. The conductivity change of SnO₂ thick-film sensors was monitored on exposure to compounds forming in smouldering and combustion processes of natural materials, e.g. paper and wood. A strong response was observed for compounds with hydroxyl groups and the highest response was obtained for phenolic structures such as methoxyphenols and benzenediols [54.29]. The response of a SnO₂ sensor element operating at 270 °C produced no sensor response to toluene, furfural and acetone [54.30]. The impedance output from the SnO₂ sensor array was measured at 300 °C in terms of conductance changes and peak frequency changes. These changes were more the more sensitive to oxygenated VOCs than other VOCs. The sensors sensitivity could also be enhanced by Pd- and Pt-doped sensors [54.31].

Considerable development work is being carried out to integrate the semiconductor sensing membrane with the interrogation and control circuitry on a silicon chip using conventional silicon technology. Using an industrial 0.8-μm complementary metal-oxide-semiconductor (CMOS) process with subsequent micromachining steps, a CO gas-sensor chip was fully integrated with advanced analogue and digital circuitry on a chip for temperature control in the range 170–300 °C with an error of ±2 °C. The whole device was encapsulated using standard microelectronic packaging methods. It was reported that the integrated sensor was capable of measuring CO at the concentration of 5 ppm with a resolution of ±0.1 ppm [54.32].

Solid-state sensor technology relies on the preparation of semiconductor films. SnO₂ sensors suffer from a lack of selectivity and drift. The high temperature re-

quired for the surface reactions encourages the growth of grains by coalescence, giving rise to unstable materials. Recent progress in the synthesis of nanomaterials has been employed to overcome some of these problems. Oxide powders were thermally evaporated under controlled conditions without the presence of a catalyst to deposit single-crystalline SnO₂ nanobelts on a platinum IDE structure on an alumina substrate. The surface-to-volume ratio was very high but the size was large enough to contain the depletion layer within the belt. SnO₂ nanobelts were sensitive to CO, NO₂ and ethanol and therefore suitable for applications in breath analysis and food control [54.33].

Advanced methods of information technology are increasingly being used to address the issues regarding selectivity and specificity. A sensor array of two commercial metal-oxide gas sensors was developed to discriminate six organic solvents over a wide concentration range of 2–200 ppm in air. The baseline drift was considerably reduced over a test period of several months by temperature cycling. Further reduction in drift was achieved by suppressing the influence of humidity via additional signal preprocessing. Relatively low computing power was required for the hierarchical pattern classification to evaluate shape features generated from the sensor response curve during temperature cycling. The technique was found to be adaptable to different operating environments, for example, to suppress false alarms from interfering gases [54.34].

Apart from SnO₂, titania (TiO₂), ferric oxide (Fe₂O₃), titania doped chromium oxide (Cr_{1.8}Ti_{0.2}O₃), molybdenum trioxide (MoO₃), tungsten oxide (WO₃), indium oxide (In₂O₃) and perovskite-type thick films (LaFeO₃) were studied for sensing a host of analytes such as H₂, CO, NO₂, H₂S, NH₃ and CH₄ in the ppb range. The characteristics of titanium dioxide nanotube hydrogen sensors were found to be completely reversible, highly selective and free from hysteresis. The sensors showed response times of approximately 150 s at a nominal operating temperature of 180 °C. The sensitivity of the nanotubes was believed to be due to chemisorption of hydrogen onto the titania surface as electron donors. Increased operating temperature improved the sensitivity significantly but reduced the response time. Samples with smaller pore diameter (46 nm) were more sensitive to hydrogen than those with larger pore diameter (76 nm). The sensor showed response to high concentrations of oxygen but the response time was long and the recovery was not complete [54.35]. Ion-beam-deposited thin films of MoO₃ on alumina substrates with gold IDEs showed selective

response to NH_3 at 450°C . Similarly prepared WO_3 thin-film samples were highly sensitive to NO_2 . Both films exhibited orthorhombic structure but the poly-

morphic dissimilarities between two types of materials were responsible for the difference in their sensing behaviour [54.36].

54.3 Macrocylic Compounds for Sensing

Phthalocyanine (**Pc**) molecules belong to a class of macrocyclic compounds. Because of their 18 π -electron aromatic macrocyclic structure these molecules exhibit important optoelectronic, photophysical, conducting and photoconducting properties [54.37]. As shown in Fig. 54.2, the **Pc** structure is closely related to the naturally occurring porphyrins, which form a core skeleton in the haemoglobin and chlorophyll. Organic molecular nanostructures have huge potential but suffer from unstable device characteristics, low resistance to adverse environments (e.g., temperature, humidity, oxygen, etc.) and lack of reproducibility of material composition, purity, and fabrication conditions. **Pcs** are, however, tinctorially strong, chemically and thermally stable. The materials are abundantly available in a highly purified form.

54.3.1 Preparation of Sensing Membranes

Sensing is essentially an interaction mechanism of a solid surface and analytes and the molecules are organised in ultrathin films, offering a large surface-to-volume ratio to improve the probability of these reactions taking place. A variety of methods such as direct-current **DC** and radio-frequency **RF** sputtering [54.38], thermal evaporation [54.39], Langmuir–Blodgett (**LB**) [54.40], spin coating [54.41] and self-assembly (**SA**) [54.42] are employed to produce organised thin organic films.

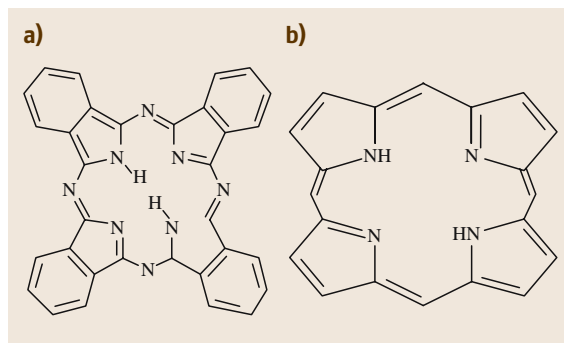


Fig. 54.2a,b Chemical structure of: (a) a metal-free phthalocyanine and (b) porphyrin molecule

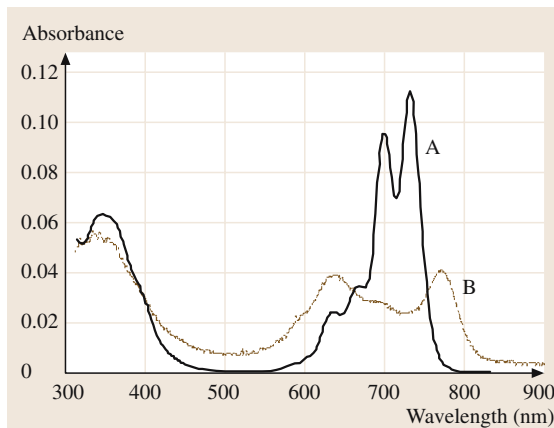


Fig. 54.3 Typical UV–visible spectra for thin films of metal-free phthalocyanine molecules in: (A) chloroform and (B) an **LB** film

The substrates play a key role in determining the film quality and the structure. It is found that the island density and coverage ratio of copper tetra-tert-butyl phthalocyanine (**CuTTBPc**) on hydrophilic glass surfaces were small during the early growth stage of vacuum deposition due to a small nucleation rate arising from

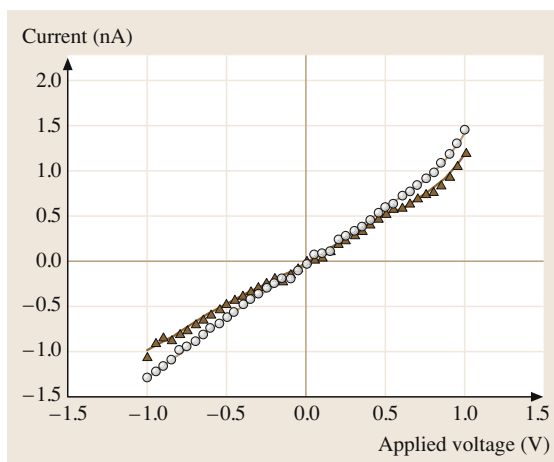


Fig. 54.4 Dark- (triangles) and photoconduction (circles) through spun **NiPc** films

the weak interaction of molecules. The film surface was also rough. The high nucleation rate of CuTTBPc on a silanized surface, on the contrary, produced a much smoother morphology consisting of densely distributed fine-grain clusters. The aggregation of the molecules took place in domains during the formation of the LB monolayer, separately distributed on the glass surface. These domains acted as active sites for the nucleation and growth of the later deposition process and high-density clusters were thus found in the early growth stage. X-ray diffraction (XRD) patterns demonstrated that the LB films grown on the glass substrate were more crystalline than those on other substrates [54.43].

The wet techniques, namely LB, spin coating and SA, are preferred to high-energy deposition because the deposition takes place at room temperature, reducing the possibilities of the chemical decomposition of the molecules caused by thermal damage. These deposition methods are considered below.

Langmuir–Blodgett (LB) technique

The LB method involves the transfer of individual Langmuir monolayers of amphiphilic molecules which self-assemble in an ultra-clean water subphase onto a suitably prepared solid substrate. An LB film, which may consist of a single or many layers, up to a depth of several visible-light wavelengths, are generally produced with fine control over the thickness and geometry to the precision at the molecular level. With the Y-

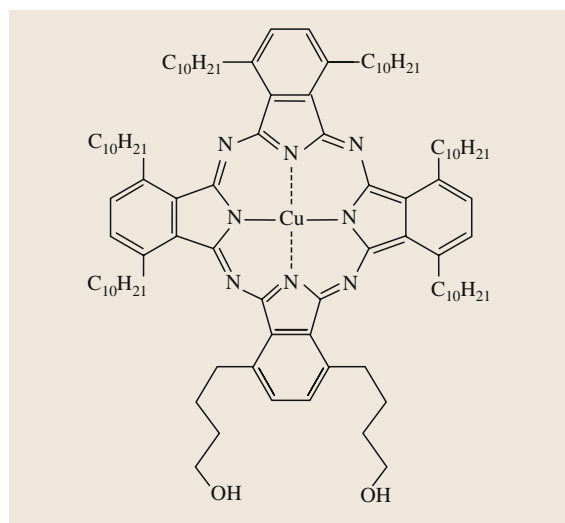


Fig. 54.5 Chemical structure of 1,4-bis(4-hydroxybutyl)-8,11,15,18,22,25-hexakisdecyl phthalocyaninato copper molecules

type construction, these LB films are essentially layered materials with alternating hydrophobic/hydrophilic regions. Non-centrosymmetric LB films can be obtained by X-type construction on immersion of the substrate from air into the water only, or Z-type deposition on withdrawal of the substrate from water to air [54.44].

Macrocylic molecules are not generally amphiphilic and are therefore not considered suitable for LB deposition. The synthesis of a series of specially designed Pc molecules for LB deposition has now been reported in the literature over the last two decades [54.45–47]. Figure 54.5 shows the chemical structure of asymmetrically substituted 1,4-bis(6-hydroxyhexyl)-8,11,15,18,22,25-hexa-octylphthalocyanine molecules belonging to this family of custom-designed molecules [54.48].

To start with LB deposition, a Langmuir monolayer is formed from a spreading solution of substituted Pc derivatives in an appropriate volatile solvent (such as chloroform) on the highly purified water subphase

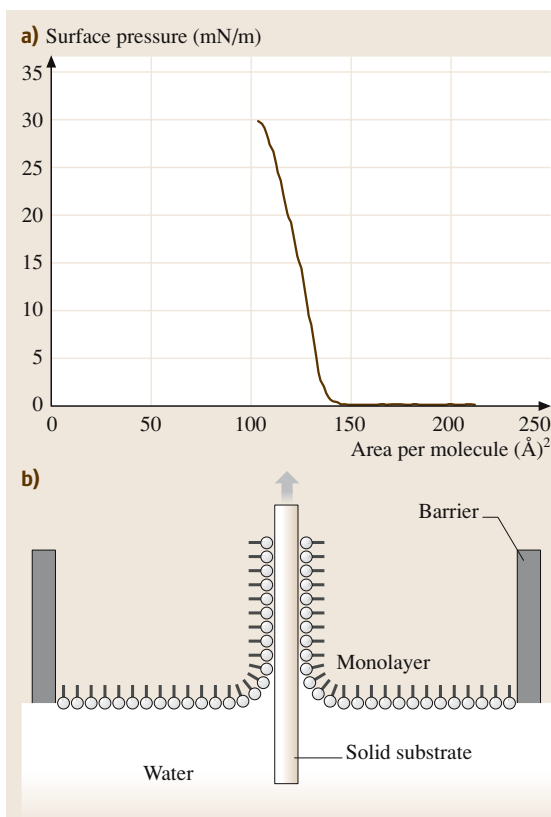


Fig. 54.6a,b Deposition of Langmuir–Blodgett film: (a) surface pressure–area isotherms and (b) transfer of floating monolayer onto a suitable substrate floating monolayer

(Milli-Q water of resistivity of 0.18 M Ω m) contained in an LB trough made of a hydrophobic material (teflon). The solvent is allowed to evaporate fully and the molecules are then slowly compressed to the desired organisation by closing a servo-controlled teflon barrier straddling the air/water interface at a constant speed (say about 100 mm/min). As shown in Fig. 54.6, the phase changes of the molecules during the compression are monitored by recording the surface pressure–area isotherm. The pressure begins to rise rapidly as the solid phase is reached while the area per molecule remains approximately constant. The molecules are transferred from the water/air interface to the substrate by dipping the substrate in at a constant speed of about 10 mm/min when the surface pressure is in the range 20–40 mNm⁻¹. However a dipping rate as high as 1000 mm/min was used to produce LB films of 5,10,15,20-tetrakis(3,4-bis[2-ethylhexyloxyphenyl])-21H,23H-porphine (EHO) molecules. Atomic force microscopy (AFM) showed a morphology of these films which is highly conducive to the host–guest interaction. The presence of isolated micron-size domains, which were themselves composed of grains of several nm in diameter, was observed [54.49].

The molecular orientation in LB films of these novel substituted Pc derivatives was investigated by spectroscopic methods, including electron spin resonance, reflection adsorption infrared spectroscopy (RAIRS), polarised infrared (IR) spectroscopy and AFM. Amphiphilic mesogenic Pcs afforded the best films. These all contained a good degree of molecular order; furthermore, the type of molecular stacking in the films was found to depend upon the length of the side chains and the central metal [54.50].

Spin-Coating Method

The spin-coating method, designed for defining patterns on silicon wafers in integrated circuit (IC) technology, has now been adapted for deposition of thin organic films on suitable substrates [54.51]. Substrates are vacuum held onto the rotating chuck of a photoresist spinner. The spread solution was prepared in the same way as done for the LB deposition and a small volume of the solution is dispensed onto the rotating substrate surface from a microsyringe held at a distance of 5 mm above the spinner platform. A spinning time of 30 s for a solution concentration of 0.5–5.0 mg/ml is found to be sufficient in order to produce a uniform, homogeneous film. The rotation speed is normally varied between 1000–6000 rpm and the speed of spin and the concentration of solution

largely determine the thickness of the resulting films. The dependence of the film thickness d on spin speed ω and time t may be written in terms of an empirical expression [54.52]:

$$d = \frac{C}{\sqrt{t}} \omega^{-x}, \quad (54.1)$$

where C depends upon the evaporation rate of the solvent, and the viscosity and density of the solution. The index x is predominantly determined by the evaporation rate of the solvent. If the evaporation of the solvent is independent of the spin speed, x takes on a value of 2/3. For most solvents used, the evaporation varies with $\sqrt{\omega}$ and x becomes 1/2. A value of $x = 1$ is a valid approximation for slow evaporation [54.53, 54].

Spin coating is a simpler and more cost-effective method than LB deposition and also there is no requirement for the molecules to be amphiphilic. The LB technique, however, provides more precise control over the film thickness [54.55]. Both spun and LB thin films of an octa-alkyl-substituted phthalocyanine molecules displayed similar characteristic features in their optical absorption spectra. The appearance of Davydov splitting and the XRD pattern indicated the same crystallinity of the films [54.56]. Different physical structure and molecular packing may result from the use of different wet techniques. For instance, asymmetrically substituted OC molecules (Fig. 54.5) were deposited by LB and spin-coating techniques. XRD patterns and UV–VIS spectra of these films indicated some degree of anisotropic ordering within the LB films compared to spun films [54.57]. The spin-coated films were, however, exceptionally smooth, even and free from crystallites. A noncompact spin-coated film of phthalocyanine molecules bearing one crown on ether ring exhibited faster response and reversal time than the well-ordered and tightly packed LB films [54.58].

Self-Assembled Monolayers (SAMs)

Self-assembled monolayers (SAMs) have attracted considerable interest because of their simplicity of preparation, stability and versatility. The introduction of due functional groups into SAMs allows changes of state and property on surface or interface. These attributes make SAM films attractive for the sensor technology [54.59, 60]. SAM films are formed via reaction of a functional group within a molecule with the surface of a solid substrate. The methoxy or trichlorosilylalkyl derivatives of molecules form weak bonds with hydroxyl groups and are suitable for the formation of a SAM on an oxide surface such as indium tin-oxide

(ITO) and glass (SiO_2) [54.61]. A monolayer layer can also be self-assembled on gold or silver substrates if the molecules are attached with thio or disulphides groups at one end [54.62]. In supramolecular aggregation during the self-assembly process, the interactions between the adsorbate molecules are as important as those between adsorbates and the surface. The size and aggregation pattern for the formation of supramolecular structures (J-aggregates) are controlled by changing the number of polar sulfonic groups of meso-tetra (4-sulfonatophenyl) porphine [54.63].

Much fundamental work on SAM films was initially undertaken with simple functionalised alkane derivatives [54.64]. In the case of alkylthiols on gold, it is well established that the monolayer films were densely packed with a high degree of order, with chains typically oriented at about 65° to the substrate surface. In recent years, work has been undertaken to include a single tether chain with thiol or disulphide derivatives for SAM formation on gold-coated glass substrates. The molecular orientation with respect to the gold surface was found to depend upon the different alkyl-connecting chain lengths and different macrocyclic peripheral side chains [54.65]. The core of substituted Pc molecules bearing mercaptoalkyl groups is not parallel to the surface if the tether chain is increasingly long. For sufficiently long tethers, the core becomes oriented almost normal to the surface [54.66]. SAM films of thiol-derivatised cobalt phthalocyanine complexes on gold electrodes was susceptible to destruction via oxidative and reductive desorption and their potential applications as a sensor were therefore limited [54.67].

Self-assembled ultrathin films have recently received considerable interest because they allow fabrication of supramolecular assemblies with tailored architecture and properties [54.68]. Multilayered films can also be built up by sequential deposition of polycations and polyanions based upon electrostatic attraction [54.69–71]. Alternate deposition of planar functional molecules meso-tetra (4-sulfonyl) porphyrin (TPPS_4) or copper phthalocyaninetetrasulfonic acid, tetrasodium salt (CuTsPc) and cationic bis (pyridinium) salt was made to produce a self-assembled multilayered film [54.72]. Multilayer assemblies of porphyrin and phthalocyanine were also prepared by alternating deposition of oppositely charged rigid planar molecules of sodium (phthalocyanine tetra-sulfonate)cobalt $\text{Na}_4[(\text{CoTsPc})]$ and tetrakis(N,N,N-trimethyl-4-anilinium) porphyrin cobalt $[(\text{CoTAP})]\text{Br}_4$ [54.73]. A further investigation of the C-11 Pc derivative SAM has shown that the monolayer film was stable, exhibiting no oxidation and

only minor orientation changes on the gold surface over a period of 24 months as determined by infrared and fluorescence spectroscopy. The long active life of the Pc SAM was thought to be due to the macrocyclic ring possibly preventing oxidation of the thiolate root [54.74].

54.3.2 Thin-Film Properties

Surface/interface composition at the nanometre scale is believed to affect the organisation and chemical/physical properties of organic thin films. This long-range organisation determines both the optical and electrical properties of thin-film materials, leading to unique, new chemical sensor platforms. The overlap of π -electron wave functions from Pc molecule to Pc molecule is likely to occur between adjacent stacks in a thin-film form, giving rise to the broadening of energy levels [54.75].

Phthalocyanine Films

Phthalocyanines can exist in different polymorphic forms: α and β are among the most common. Both forms possess the herringbone structure for the stacking arrangement. The angle between the stacking axis and the normal to the molecular plane depends upon the method of preparation, types of phthalocyanine (metallo or metal-free) and substituents. For example, this angle is generally 25° and 45° for the α and β phases, respectively [54.76]. A typical absorption spectrum in the UV–visible range is shown in Fig. 54.3 for a substituted metal-free phthalocyanine (HPC) deposited on a glass substrate (similar to one shown in Fig. 54.2). The Q absorption band of the LB film was broader than one for the chloroform solution and appeared at the wavelengths between 650–800 nm. The Davydov splitting became apparent when compared with the spectra for the molecules in chloroform. The film forming properties and the stack order were both found to be influenced by the chain length of the ring substituents [54.77, 78]. An analysis into the in-plane molecular arrangements in the LB films of Pc compounds with short and long alkyl chains was performed by decomposing the absorption spectra into Gaussian–Lorentzian components. Long-chain molecules were found to be organised in a herringbone structure leading to the Davydov splitting of the Q-band. For short-chain molecules, the largely broadened blue component of the Davydov doublets was suppressed possibly due to the domination of an anisotropic stack-like molecular arrangement in LB films [54.79]. Characteristic Q-bands observed at about 630 nm in the UV–VIS spectra of the self-assembling monolayers (SAMs) of 2,9,16-Tri(tert-butyl)-23-(10-

mercaptodecyloxy)phthalocyanine and its disulfide on gold substrates were broadened and blue-shifted relative to those observed in solution [54.80]. The formation of high-quality Langmuir–Blodgett films of copper and nickel phthalocyanine derivatives substituted with short branched chains is reported. Ellipsometric and polarised optical absorption measurements suggest that the phthalocyanine molecules have a preferred orientation, with their large faces perpendicular to the dipping direction and to the substrate plane [54.81].

Electrical properties of **Pc** molecules are very dependent upon the doping and the nature of the substituents in the macrocyclic. Copper(II) tetrasubstituted phthalocyanine complexes, **CuPcX** (where $X = -\text{NO}_2$, $-\text{NH}_2$, $-\text{SO}_3\text{H}$, and $-\text{OH}$) exhibited an improvement of conductivity by five orders of magnitude over its unsubstituted counterpart. Further enhancement in conductivity was achieved by iodine doping. This increase in conductivity was attributed to the possible decrease in the metal–metal bond distance [54.82]. The effect of oxygen on charge transport was investigated by measuring the time-dependence behaviour of the film conductivity of zinc **Pc** (**ZnPc**) layers at a stable pressure of about 0.1 mbar of air and it was found that the conductivity gradually decreased by an order of magnitude with time. The conductivity reached saturation within several minutes for the film thickness smaller than 500 nm whereas similar electrical degradation in a 7- μm -thick film required more than half an hour. This thickness-dependent decrease in conductivity was caused by the velocity of the oxygen out-diffusion processes in the films [54.83]. The effect of doping on conductivity of a thermally deposited composite film of α -nickel phthalocyanine (**NiPc**) and the strong electron acceptor tetracyanoquinodimethane (**TCNQ**) was investigated by exposure to air. The doping took place at a much faster rate in the composite than in **NiPc** films. Values of 1.3 eV and 1.9 eV for the activation energy were obtained for devices incorporating **TCNQ** and the non-**TCNQ** devices, respectively, and the conductivity of the composite film was therefore found to be 20 times larger than the **NiPc** film [54.84]. Due to the excellent ability to photogenerate free charge carriers, the conductivity of the metal phthalocyanine (**MPc**) film is increased considerably as it can be seen from Fig. 54.4 showing the dark and illuminated current–voltage characteristics of a 10-nm-thick spun nickel phthalocyanine (**NiPc**) sandwiched between two metal electrodes.

Phthalocyanines of rare-earth metals primarily occur in the form of bisphthalocyanines with a sandwich-type structure and the cation is eightfold coordinated to the

two macrocycles. One of these phthalocyanine rings is virtually planar while the other is significantly distorted [54.85]. The Q-band in the absorption spectra of spun films of substituted lutetium bisphthalocyanine molecules becomes broader and red-shifted compared to the molecules in chloroform solution. Heat treatment of the film at temperature above 120 °C introduced molecular ordering, possibly due to the edge-to-edge interaction between neighbouring **Pc** moieties [54.86]. All these factors have led to successful formulation of phthalocyanines as chemically sensitive membranes for environmental monitoring [54.87–91].

Porphyrin Films

Inherent stability, unique optical properties and synthetic versatility of porphyrins and metalloporphyrins are exploited in sensor applications. Synthetic porphyrins and metalloporphyrins incorporated into polymers, glasses and Langmuir–Blodgett films matrices are used for detection of NO , CO_2 and O_2 because of the ability of heme to bind these gases. Metalloporphyrin sensing arrays have been used for the detection of organic odourants, such as amines, thiols and phosphines [54.92]. Porphyrins also combine readily with metals, coordinating with them in the central cavity. Many gas sensors take advantage of analyte binding to the centre metal atom in the porphyrin ring and a detectable optical change occurs as a result. In particular sensing properties of iron, zinc, copper, nickel, and cobalt containing porphyrins have now been studied [54.93–96]. The porphyrin rings of a thiol-derivatised cobalt (II) 5,10,15,20-tetrakis (4-tert-butylphenyl)-porphyrin (**CoTBPP**) molecules were immobilised on the surface of a 4-aminothiophenol (4-ATP) self-assembled monolayer (**SAM**) on gold (111) by in situ axial ligation. The reconstruction of the herringbone structure took place probably due to adsorption/desorption processes of molecules [54.97].

Substituents have considerable effects on the structure of porphyrin monolayers; for example the interaction between the porphyrins deposited on the **ITO** substrate without bulky tert-butyl groups was much stronger than that of the porphyrins with bulky tert-butyl groups [54.98]. Hybrid molecule-silicon capacitors were formed by attaching [5-(4-dihydroxyphosphorylphenyl)-10,15,20-trimesitylporphinatozinc (II)] porphyrin complexes to silicon oxide via a phosphonate linkage and the presence of multiple distinct peaks in electrical characteristics were associated with oxidation and reduction of the molecular monolayer [54.99]. Multilayered thin films of

porphyrins bearing bulky alkoxyphenyl substituents at two of the four meso-positions and phenyl phosphonates at the other two were deposited on the ITO substrates using the zirconium phosphonate linkage. Molecular aggregation was prevented by the addition of sterically demanding 2,6-di(n-hexoxy)phenyl substituents to the meso-positions of the porphyrin skeleton. The com-

pounds were able to form thin films without the presence of large molecular chromophore interaction relative to sterically unhindered porphyrins [54.100]. Highly conducting tetra-ruthenated cobaltporphyrin (Co (TRP)) and tetrasulfonated zincporphyrin (Zn (TPPS)) multi-bilayer films were prepared by electrostatically assembly for sensor applications [54.101]

54.4 Sensing with Phthalocyanine and Porphyrin

The interaction of reactive gases such as O₂, NO₂ and NH₃ with the phthalocyanine group of compounds has been shown to result in conductivity changes of these compounds. This effect was associated with charge transfer between the electron-donating phthalocyanine molecule and the electron-accepting gas [54.102].

54.4.1 Amperometric Sensor

The action of the oxidising gases O₂ and NO₂ and the reducing gases NH₃ and H₂ diluted in N₂ on the electrical conductivity of thin layers of zinc hexadecafluorophthalocyanine (ZnF₁₆Pc) and zinc phthalocyanine (ZnPc) was investigated. The conductivity ZnF₁₆Pc increased in the presence of NH₃ whereas ZnPc exhibits no sensitivity to this gas. The sensitivity to H₂ was higher for ZnF₁₆Pc than for ZnPc. This behaviour was related to the redox potentials of the two phthalocyanines [54.103, 104].

Response and recovery times for LB CuTTBPc films on platinum IDE are shown in Fig. 54.7 for different gas concentrations of 1–5 ppm. These times were determined from a model based upon the kinematics of adsorption and desorption of NO₂ gas molecules. It was found that recovery times were longer than response times, irrespective of gas concentration. The response time constant varied in the range 105–159 s. Repeated exposures to NO₂–air mixtures resulted in gradual degradation of reproducibility. The response was also believed to be influenced by the past history of the exposure of the sensor to the NO₂ gas, if the recovery time constant was longer than the switch-off time. These factors show that CuTTBPc is a material that is suitable for fabricating a disposable *single-shot* NO₂ sensor for practical uses [54.105]. LB films of an asymmetrically substituted phthalocyanine, 1,4-bis (4-hydroxybutyl) 8,11,15,18,22,25-hexaethylphthalocyanine molecules showed selective responses to 1–5 ppm concentrations of NO₂ at the room

temperature. Moisture had no effect on the response because of the hydrophobic nature of the alkyl groups. Exposure to Cl₂, NH₃, SO₂ and CO at ppm levels produced no measurable response [54.106]. Further work on the LB films showed that the NO₂ response was larger for the LB film produced by dipping the substrate perpendicular to the electrode finger than those dipped parallel to the fingers. This preferential behaviour was attributed to the ease of conduction along the long chains [54.107]. LB films of meso,meso'-buta-1,3-diyne-bridged Cu(II) octaethylporphyrin dimer molecules were used to fabricate amperometric sensors operating temperature of 90 °C. LB multilayers were reported to be selectively sensitive to small concentrations of NO in air without being affected by the contemporary presence in NO₂, CH₄, C₂H₅OH and CO [54.108].

Thickness, operating temperature, doping and post-deposition treatment of sensing membranes all affect the sensor performance. Vacuum-deposited CuPc of thickness 50–400 nm onto gold IDEs on glass substrate were

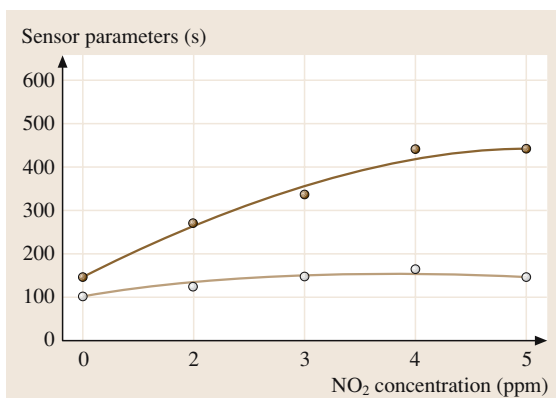


Fig. 54.7 Response (broken line) and recovery (solid line) of a 15-layer LB film of CuTTBPc when exposed to NO₂ periodically at 2-min intervals at 1, 2, 3, 4 and 5 ppm levels in succession

exposed to NO₂ gas of concentration 10–100 ppm. The film thickness had little effects on the room-temperature sensitivity, but the recovery ratio decreased with increasing film thickness. Thicker films showed improved recovery at 100 °C [54.109]. Due to the tetra-tert-butyl substitution on the periphery of phthalocyanine, CuTTBPc films have a larger lattice spacing, higher film resistance. Sublimed film CuTTBPc film showed a higher recovery ratio in the NO₂ sensing experiments [54.110].

NO₂ and O₂ treatment of evaporated CuPc films that were 160–200 nm thick on IDE improved the sensitivity and recovery time when exposed to 0.5–5 ppm of NO₂ gas. A Langmuir behaviour was predicted for the adsorption/desorption processes on the weakly binding surface sites of NO₂-treated CuPc films having a single activation energy [54.111]. When the CuPc films were subjected to post-deposition treatment by cooling down to 77 K, the mechanism of NO₂ adsorption, however, followed the Elovich model [54.112]. 100-nm-thick CuPc and iron Pc (FePc) films were heat treated at 240 °C under vacuum and then the responses of the films to the exposure of NO₂ in the 8–12.8 ppm range were monitored. The CuPc films were transformed to a stable β -phase and the adsorption occurred on the ligand rings. NO₂ was, on the other hand, adsorbed on both the ligand ring and the central ion of the FePc films [54.113]. An amperometric sensor was fabricated using the spun film of tetra-iso-propoxyphthalocyaninato copper (II) (i-Pro-CuPc), and the response, recovery and cyclic properties of the sensor were very temperature sensitive. The results showed ideal gas sensitivity at 65 °C and a good linear relationship existed between the resistance of gas sensor and the concentration of NO₂ [54.114]. 25-nm-thick NiPc films were deposited between the source and the drain of a field-effect transistor (FET) and the change in the drain current exhibited a linear dependence on the ozone concentration in air, in the range 0–150 ppb [54.115, 116].

Water-soluble copper phthalocyanine tetrasulfonate (CuTsPc) molecules were alternatively self-assembled with a bipolar pyridine salt on a 3-mercaptopropionic-acid-modified Au electrode to produce copper ion-selective electrodes (ISE). These ISE were found to have low resistance, short conditioning time and fast response. The electrode potentiometric response showed selective potentiometric response to Cu²⁺ ion in the range from 1×10^{-5} – 1×10^{-1} M, independent of the pH of the solution between pH 1–5. The response was dependent on the nature of the medium and the detection limit was found to be 7×10^{-6} M in acetate buffer

(0.2 M, pH 4.5) [54.117]. Carbohydrates were detected by flow-injection analysis at the water-soluble cobalt phthalocyanine tetrasulfonate (CoTsPc)-modified electrode with high sensitivity. Detection limits obtained in this manner range from 150 pmol injected for glucose and fructose to 600 pmol injected for maltose and sucrose [54.118].

54.4.2 Optical Sensors

The amperometric approach generally suffers from long response times and their use is restricted due to possible fire and explosion hazards caused by electrical power. Organic systems are also prone to slow degradation arising from the chemical reactivity of charge carriers. Optical transduction techniques, on the other hand, offer high speed, high precision, immunity to interference, and remote sensing capabilities for detection of trace gas species.

Optical Absorption and Fluorescence Spectra

Optical transmittance of thermally evaporated lead phthalocyanine (PbPc) and nickel phthalocyanine (NiPc) thin films changed on exposure to NO₂ in the IR region of the optical absorption spectrum [54.119]. The disappearance of Davydov splitting from the UV–VIS absorption spectra of 20-layer-thick LB films of 1,4-bis(4-hydroxybutyl)-8,11,15,18,22,25-hexakisdecyl phthalocyaninato copper molecules was observed when

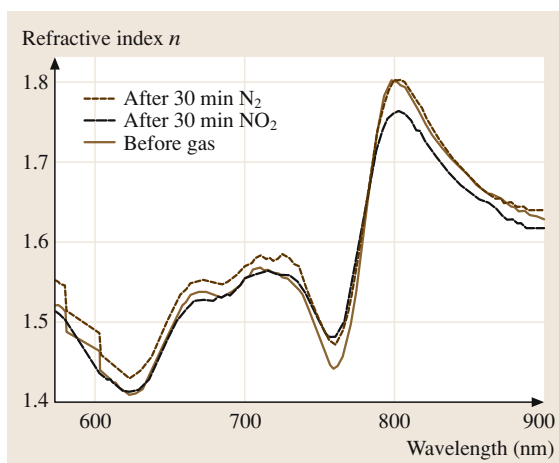


Fig. 54.8 The dependence of the refractive index of octa-substituted phthalocyanine on the wavelength before and after exposure to 90-ppm NO₂ gas for 30 min. The recovery was examined after flushing of the LB films with N₂ for 30 min

the films were exposed to 100 ppm of NO₂ [54.120]. The local interaction between the central metal atoms and NO₂ distorted the herringbone structure. Gas adsorption and desorption processes were described by exponential functions, both characterised by a single time constant. The dispersion relation in Fig. 54.8 for the refractive index was obtained by using simple Kronig–Krammer transform of absorption spectra of the LB thin film of another octa-substituted metal-free phthalocyanine molecules. The full recovery of the refractive index to its original value was achieved after flushing with the N₂ gas for wavelengths larger than 782 nm. Ellipsometry measurements on thin films of tetrasulfonated copper phthalocyanine showed increases in the real and the imaginary parts of the dielectric constant on exposure of NO₂ gas possibly due to the rise in the density of dipoles and higher absorption, respectively [54.121].

LB films of 5,15-bis(4-aminophenyl)-10,20-bis[3,4-bis(2-ethylhexyloxy)phenyl]-21H,23H-porphine were prepared at a dipping rate as high as 500 mm/min and an inhomogeneous structure was obtained with high porosity. These porphyrine films showed much faster response to 4.6 ppm of NO₂ gas and a sensitivity of 60% relative absorbance change at 439 nm was achieved. Full and accelerated recovery of the original spectrum was also achieved by gentle heating at 353 K. The sensor was reported to have a self-life as long as one year [54.122]. Kinetic UV–VIS spectroscopic measurements were made on a porphyrin-entrapped sol–gel film for NO₂ sensing. The method was capable of detecting levels as low as 176 ppb of NO₂ at room temperature and showed a fast and fully recovery [54.123].

The blue colour of rhodium phthalocyanine LB films becomes transparent upon chlorination. A pronounced quenching of the characteristic triplet centred on the Q-absorption band took place at 662 nm, but the absorption band in the near UV part of the spectrum remained almost unaffected. Optical spectra partially recovered when the exposed film was left in air for several hours. Overall effects of chlorination (quenching and recovery) depended on the period of exposure [54.124]. LB gadolinium phthalocyanine films were exposed to 10 ppm of chlorine and changes in optical absorption spectra were observed as a result of the oxidation of the film by chlorine to form [Gd(Pc)₂]⁺ species. The kinetics of such changes were described by a polyfunctional process with different decay time constants and the effects were found to be highly reversible [54.125].

SAM films are of interest because the chemically bound monolayer is intrinsically more robust than LB

films and the response times are also expected to be faster. Monolayers of diphtalocyanine disulfide Pc molecules were self-assembled on gold-coated glass substrates with the Pc macrocycle oriented parallel to the metal surface. Changes of the reflectivity signal were observed on exposure to NO₂ gas in proportion to the concentration [54.126]. The fluorescence emission spectrum of each of the Pc SAMs was obtained by exciting each monolayer along the longitudinal axis of the optical waveguide via laser-induced evanescent-wave stimulation. The use of the longer mercaptoalkyl connecting chain appeared to inhibit quenching of the electronically excited state through energy transfer to the metal layer. Fluorescence response to 10 ppm NO₂ was found to be selective with no interference from CO and CO₂ [54.127].

Surface Plasmon Resonance Technique

The surface plasmon resonance (SPR) technique is based on the excitation of surface electromagnetic waves of transverse magnetic (TM) modes travelling along the interface between a metal and a dielectric medium. Using a semicylindrical prism of high refractive index n_p in Kretschmann's configuration, as shown in Fig. 54.9, these waves are excited by a p-polarised light of wavelength λ via the evanescent field [54.128]. The evanescent field will be completely attenuated at a particular angle of incidence corresponding to resonance and there will be no reflection. Figure 54.10 shows the SPR results when a spun film of nickel(II) tetrakis(4-cumylphenoxy)phthalocyanine (NiPc) molecules was subjected a sequence of exposure to ozone (O₃) at 2 ppm. The period of 20 mins was allowed for the film to recover to its original state. The shift of the resonance angle $\Delta\theta_a$ was believed to be caused by a change in the refractive index of the film due to the interaction of O₃. $\Delta\theta_a$ is

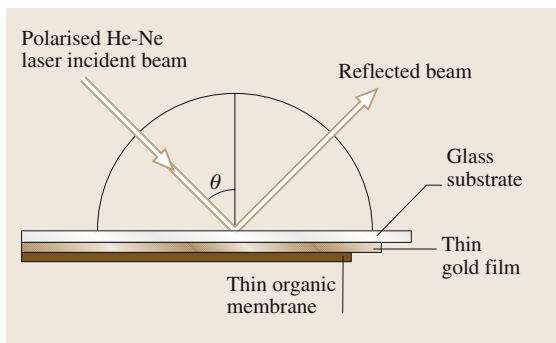


Fig. 54.9 A Kretschmann-type configuration to measure SPR data of CA thin films

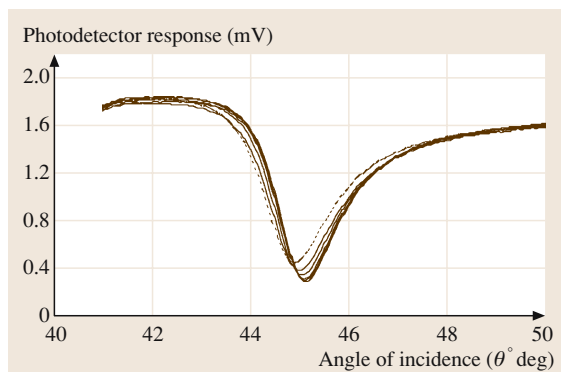


Fig. 54.10 The effect of a sequence of exposure at 20-min intervals of 2-ppm O_3 on a NiPc spun film. The broken line represents the system prior to exposure

usually written in the form [54.129]:

$$\Delta\theta_a = \frac{(2\pi/\lambda)(|\varepsilon_m|\varepsilon_i)^{3/2}d}{n_p \cos\theta(|\varepsilon_m| - \varepsilon_i)^2\varepsilon}(\varepsilon - \varepsilon_i), \quad (54.2)$$

where $|\varepsilon_m|$ is the modulus of the real part of the dielectric constant of the gold film; ε_i is the dielectric constant of the medium (air in this case) in contact with the organic thin layer of thickness d and the dielectric constant ε . The resonance minimum moved to an increasingly larger angle after each exposure to ozone. The half-width became narrower at the same time and this was believed to be the consequence of oxidation of the phthalocyanine ring by ozone, which resulted in bleaching of the dye. SPR responses are in general slower and less readily reversible than semiconductivity changes but the technique is capable of providing additional information on the interaction between the gas and films as the shifts in both the angle and depth of the resonance can be monitored. Experimental data from the SPR response of CuPc films to nitrogen dioxide were input to a design process using an evolutionary algorithm. 25-nm-thick gold layers were found to exhibit 42.9% greater contrast than layers with a thickness of 50 nm. Using evolutionary design predictions, further modifications could be tested for available materials, and redundant layers might be eliminated. By inclusion of the external optics, a design could be selected to accommodate poor precision ($\pm 0.5^\circ$) in the incident angle and a possible multi-layer solution was shown using Teflon AF 1600, with a refractive index similar to 1.3. The predicted NO_2 response showed an improvement compared with the classical SPR configuration, and the incident angle chosen by the genetic algorithm for the interrogation of these

layers was close to a stationary point in the absolute response curve [54.130].

Figure 54.11 shows the transient response when thick films of copper hexadecafluorophthalocyanine ($CuPcF_{16}$) were exposed to 100 ppm of NH_3 for 2 min followed by the injection of dry air for a further 2-min period. The film was exposed to 200 ppm of NH_3 for the next cycle. The SPR response was very rapid and the recovery was complete. Unsubstituted CuPc was practically insensitive to this gas. Strong electron-withdrawing fluorine substituents made this substituted compound sensitive to reducing gases such as NH_3 [54.131].

SPR experiments were performed on films of different phthalocyanine derivatives in order to investigate the interaction mechanism with NO_2 . SPR signal concentration as low as 1 ppm was measured using spun films of crown-ether-substituted phthalocyanines and octa-3,7,11-trimethyldodecyloxy phthalocyanine [54.132]. Adsorption of NO_2 took place both on the surface of the film and into the bulk. The surface adsorption increased the film thickness while the change in the real part of the film refractive index occurred due to the diffusion of NO_2 into the LB film. These processes were reversible and had no effect on the optical absorbance of the LB film. The SPR response of the bare Ag film to the exposure of NO_x (the equilibrium mixture of nitrogen dioxide and dinitrogen tetroxide) was slightly slower than the LB films of tetra-4-tert-butyl-phthalocyanitosilicon dichloride (ttb-PcSiCl₂/Ag) system. This effect was associated with the high affinity of the phthalocyanine to NO_x

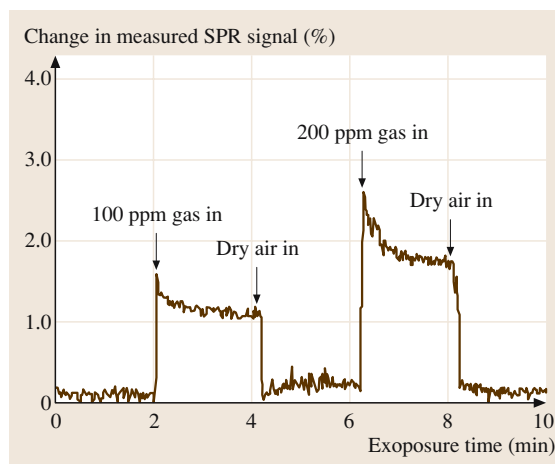


Fig. 54.11 Kinetics of the SPR response of copper hexadecafluorophthalocyanine ($CuPcF_{16}$) film to NH_3 gas at 100 ppm and 200 ppm. Measurements were at a fixed angle of $\theta^* = 46^\circ$

gas. The NO₂ gas was also able to reach the Ag/Pc interface film and the reaction with the metal film resulted in the growth of a surface layer. This irreversible effect was found to be predominant during the recovery cycle and considerably reduced by the presence of a buffer LB layer of ω -tricosenoic (ω -TA) acid [54.133]. SPR studies showed that exposure of thin films of 18-crown-6 metal-free phthalocyanine molecules to NO₂ decreased the film absorbance. The optical permittivity of the NO₂-treated Pc over-layer was regarded as being an average of the optical permittivities of both components since the optical fields associated with surface plasmon polaritons sampled both the Pc and the NO₂ molecules. The disruption in the charge conjugation of the phthalocyanine molecules was observed because of the generation of organic radical cations as end products of the charge transfer during the physisorption process [54.134].

Using the excitation at wavelengths of 488 and 632.8 nm, SPR measurements were made on LB films of porphyrine molecules of EHO type deposited on an Ag-coated glass substrate. The resonance shift observed was larger at 488 nm than at 632.8 nm when the EHO films were exposed to NO₂ gas. The thickness and the complex dielectric constants at 632.8 nm were attributed to the island structure of the EHO LB films but the derived properties at 488 nm were believed to be caused by dispersion due to optical absorption [54.135]. Similar results were further obtained from UV-VIS absorption in these films [54.49].

54.4.3 Detection of Volatile Organic Vapour Compounds

Phthalocyanine molecules have also been extensively investigated for the recognition of volatile organic compounds (VOC). Langmuir-Blodgett (LB) and evaporated films of lutetium bisphthalocyanine (LuPc₂) molecules were deposited on indium tin-oxide (ITO) interdigitated electrodes in order to study the changes in their conductivity at room temperature due to the presence of a variety of organic vapours with different chemical functionalities. The results demonstrated the viability of the phthalocyanine thin films as the active species for systems specifically designed for the monitoring of aromatic components in food. The majority of the work on vapour sensing has, however, been based upon quartz crystal microbalance (QCM) transduction techniques. The frequency shift Δf from the nominal resonance frequency f_0 of the crystal due to the change in the mass Δm due to the adsorption of the vapours

is [54.136]:

$$\Delta m(\text{mg}/\text{cm}^2) = \frac{\Delta f(\text{Hz})}{2.26 \times 10^{-6} f_0^2}. \quad (54.3)$$

It was demonstrated that monomeric soluble transition-metal phthalocyanines R₄PcM (R=tert-butyl or 2,2-dimethyl-3-phenyl-propoxy) used as sensitive coatings for quartz microbalance transducers show reversible interaction and high sensitivity for organic solvents with high boiling point [54.137]. Quartz microbalance devices coated with LB films of copper(II) tetra-(tert-butyl)-5,10,15,20-tetraazaporphyrin exhibit good sensing response to the vapour of benzene and toluene at room temperature [54.138]. Soluble tetrakis-hexyl- and dodecylthiophthalocyaninato nickel(II) and the corresponding Pd(II) and Ag(I) complexes were investigated as sensitive materials for the detection of organic solvent vapours using quartz microbalance and interdigitated capacitance transducers. Sensor responses were found to be reversible at room temperature with response times on the order of several seconds depending on the partition coefficients of the organic solvents in the phthalocyanine film [54.139]. LB films of lanthanide diphthalocyanines such as praseodymium, ytterbium as well as octa-tert-butyl praseodymium diphthalocyanines were known to display spectroscopic changes when exposed to tobacco smoke [54.140]. The sensitivity and partition coefficient for ethanol, dichloromethane, acetone, and *n*-hexane were found to be larger for the QCM sensor using thermally annealed ordered membranes of octa(13,17-dioxanonacosane-15-sulfanyl)-substituted mesomorphic nickel(II) phthalocyanine molecules than as-coated untreated films. It was found from Raman spectroscopy that molecules with saturated C—C bonds such as ethanol interact with phthalocyanine films predominantly by the formation of hydrogen bonds and the sensor response to π -bond-containing compounds such as acetone is the result of their π - π interaction with the conjugated phthalocyanine ring [54.141].

Changes in refractive index of lutetium bisphthalocyanine (LuPc₂) in the presence of ethanol, hexanal, *n*-butyl acetate and acetic acid were successfully exploited to produce an optical-fibre sensor operating at the wavelength of 1310 nm, capable of monitoring changes up to 10 dB in the reflected optical power for the detection of an 88-mmol/l concentration of acetic acid [54.142]. UV-VIS spectra of spin-coated layers of 29H,31H-(2,4-di-*t*-amylphenoxy)phthalocyanine, Zn(II) tetra-4-(2,4-di-*t*-amylphenoxy)phthalocyanine and Zn(II) tris-(2,4-di-*t*-amylphenoxy)-[4-(4-mercapto

-phenylimino-methyl)-phenoxy] phthalocyanine were found to be sensitive to tert-butylamine, diethylamine, dibutylamine, 2-butanone and acetic acid. The selectivity was determined by both the metal and the peripheral substituents [54.143]. The synthesis of specially designed bisphthalocyanine derivatives for VOC detection has been reported and their UV spectra are selectively sensitive to vapours depending on both the metal and the peripheral substituents. This observation has led to the fabrication of an *electronic optical nose* using an array of as-manufactured sensors for the analysis of some volatile organic compounds (VOC) that are of interest in food analysis [54.144].

The electrical conductivity of sol-gel-derived spun hybrid cobalt porphyrin-SnO₂ thin films showed fast and reversible response to methanol vapours; the highest responses were observed at 250 °C. The porphyrin was thermally stable up to 300 °C within the SnO₂ ma-

trix and enhanced the methanol detection sensitivity at lower working temperatures. The incorporation of porphyrin into SnO₂ film had no effect on the detection of CO [54.145]. Hybrid thin films were produced using porphyrin and phthalocyanine and the response of absorption bands of the blend systems to the presence of VOCs was different from that obtained with a single compound [54.146].

Toluene vapour sensing has been successfully demonstrated using LB films prepared from copper tetrakis-(3,3-dimethyl-1-butoxycarbonyl) phthalocyanine (CuPcBC), copper tetrakis-(3,3-dimethyl-1-neopentoxycarbonyl) phthalocyanine (CuPcNC) and nickel tetrakis-(3,3-dimethyl-1-butoxycarbonyl) phthalocyanine molecules. Exposure to toluene resulted in a partially reversible shift in the resonance depth and position of the SPR curves and toluene could be detected down to at least 50 ppm [54.147].

54.5 Polymeric Materials

A polymer comprises of repeating molecules with the same chemical structure. The electrical conductivities of the intrinsically conducting polymer systems now range from 10^{-10} – 10^{-5} S/cm. The common electronic feature of pristine conducting polymers is the π -conjugated system, formed by the overlap of carbon p_z orbitals and alternating carbon-carbon bond lengths. Doping of the polymers can increase the conductivity to as high as 10^{+4} S/cm, a value comparable to that of a metal.

54.5.1 Conducting Polymers

Figure 54.12 shows the chemical structure of a few commonly used conducting polymers such as polyaniline, polypyrrole, and polythiophene for chemical sensors. In addition to the deposition techniques described earlier, conducting sensing filaments were produced using the scribbling and platinum-wire techniques [54.148]. Sensing membranes were also formulated by laser-induced chemical vapour polymerisation [54.149].

LB films of stearic acid, a polyaniline oligomer (16-mer) and polypyrrole were deposited on the IDEs. Composite films of polyaniline and polypyrrole mixed with stearic acid in equal ratio by weight were also deposited on similar electrodes. Different brands of min-

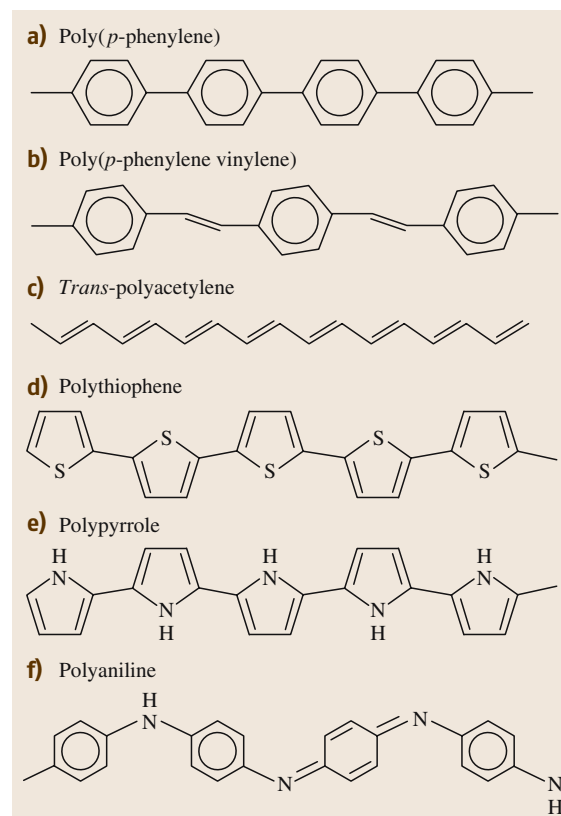


Fig. 54.12 Chemical structures of typical conducting polymers for sensors

eral water, tea and coffee were sensed by impedance measurements on the films over the frequency range of 20–10⁵ Hz. The sensor arrays composed of these films were able to differentiate tastants below the human detection threshold [54.150]. The Brunauer, Emmett and Teller (BET) surface area for an amperometric polyaniline/Au/Nafion[®] sensor prepared by the cyclic voltammetry (CV) method was found to be higher than structures prepared by the constant current (CC) method; the maximum sensitivity of CV-fabricated NO₂ gas sensors was 3.04 μ A/ppm [54.151].

The type of doping is known to influence conducting-polymer-based sensors. Composite polyaniline emeraldine salts doped with an inorganic metal complex dopant, chromium(III) trioxalate (CTO) were produced by the chemically oxidative polymerisation method. Polyanilines were also electrochemically doped with naphthalene-1,5-disulphonic acid (NSA). The conductivities of these materials were monitored by using the four-probe method when pellets of these materials were exposed to several saturated vapours of organic solvents for 30 min; the results are summarised in Fig. 54.13. The conductivity of the CTO-doped pellets was reduced by 18% on exposure to CH₂Cl₂ whereas the exposure had no effect on NSA-doped material. The response of undoped composite polyaniline salt to CH₂Cl₂ was also very small. NSA-doped pellets, on other hand, showed a significant change in conductiv-

ity on exposure to CCl₄, while only very small changes were observed for the two other polymers. The responses to dimethylsulfoxide (DMSO) and N-methylpyrrolidone (NMP) were consistently higher for all three materials than the responses to the solvents CH₃CN and methanol (MeOH). It became apparent that the interaction of solvents with the polyaniline salt became stronger with increasing polarity of the solvents. The response was reversible for all three materials. The nature of the dopant in the material and the consequent change in polarity, especially at the dopant site, and the resultant effect on π -electron delocalisation and conductivity are primarily responsible for the characteristic sensing ability of the materials [54.152]. Extruded polymer blends containing polyaniline doped with dodecyl benzene sulfonic acid were dispersed in polystyrene (PSt) matrix filaments and the electrical conductivity of these filaments increased by a few orders of magnitude when exposed to a homologous series of alcohols, including methanol, ethanol and 1-propanol. The effect was reproducible and recovery was satisfactory. The sorption of analyte molecules was believed to facilitate conduction through the insulating PSt moieties by enhancing charge-carrier mobility through hopping processes between adjacent polyaniline particles [54.153].

Multilayered thin films of polypyrrole were prepared by modifying a preformed Langmuir–Blodgett (LB) film of stearic acid by exposure to gaseous reactants such as hydrochloric acid and pyrrole vapours. A change in the electrical resistance of the film was observed when the film was exposed to polar molecules such as methanol, ethanol and acetone, while the effect of nonpolar molecules on the resistivity of the film was found to be smaller. The response of LB films to exposure to ethanol was faster than that of electrochemically deposited polypyrrole, possibly due to the high ratio of surface area to bulk volume in the LB film [54.154]. Changes in mass and conductivity of polypyrrole thin films were measured to sense the primary alcohols [54.155].

Attempts have been made to integrate polyaniline optical sensor elements with optical communication technology for remote sensing by using 1300 nm as the operating wavelength. A 1% increase in transmission was observed at 1300 nm for electrochemically prepared polyaniline films when exposed to gaseous ammonia levels as low as 6 ppm at 50% relative humidity (RH). Different concentrations of ammonia were read by calibrating the sensors at intervals of 15 s. The response time was found to be insensitive to relative humidity variations in the range 30–70% [54.156].

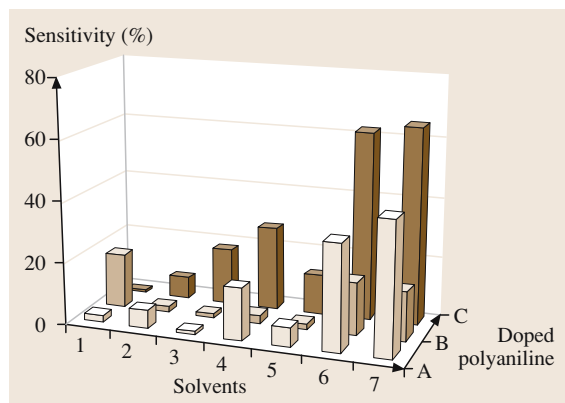


Fig. 54.13 Effects of vapour of: (1) dichloromethane (CH₂Cl₂), (2) chloroform (CHCl₃), (3) carbon tetrachloride (CCl₄), (4) acetonitrile (CH₃CN), (5) methanol (MeOH), (6) dimethylsulfoxide (DMSO), and (7) N-methylpyrrolidone (NMP) on the conductivity of: (A) pristine polyaniline emeraldine salt, and the salt doped with (B) chromium(III) trioxalate (CTO), and (C) naphthalene-1,5-disulphonic acid (NSA)

Certain specially prepared polyaniline films showed partially reversible absorbance changes in the wavelength range 620–728 nm when exposed to 50–100 ppm of ozone at room temperature. The maximum sensitivity was attained at 620 nm. The cost-effective commercial exploitation of this effect was possible since the wavelength range was compatible with implementation on low-cost plastic optical fibres and small light sources [54.157].

Efforts have been invested in the use of information technology techniques to produce smart sensors. The sensor system, consisting of an acoustic two-port resonator operating at 433.92 and 380.8 MHz was configured as a frequency oscillator, including an integrated electronic module. The polymeric membranes were tested at room temperature for response to NO₂, NO, NH₃, CO, CH₄, SO₂, and H₂S in N₂. Responses to relative humidity (RH), and organic vapours (ethanol, acetone, and ethyl acetate) were also monitored. The sensitivity was found to be high. Principal component analysis (PCA) was performed to distinguish between different vapours of low concentrations [54.158]. Fluorescence emission from polymer-immobilised dye molecules on the multi-fibre tips was studied on exposure to organic vapours. Temporal responses were found to depend upon chemical nature (for example, polarity, shape and size) of both the vapour and the polymer; these were used as input signals to train a neural network for vapour recognition. The system was able to identify individual vapours at different concentrations accurately [54.159]. Similar work was reported using the solvatochromic dye, Nile red. The dye was immobilised within various polymers. The substrate played a role in determining the sensitivity of the sensor. The sensitivity attained with microstructure glass (MSG) substrates was seven times greater than that with SU-8 photoresist-coated glass substrate and a 50% faster recovery was also achieved with the MSG substrates. The MSG sensor array was able to fingerprint the response for separate analytes with a high degree of repeatability. Using pattern-recognition techniques, sensor arrays were adaptable for gas identification and discrimination ([54.160]).

54.5.2 Ion Sensing

Water purification requires calibration of metal-ion contents. Polymeric ion-selective electrodes (ISEs) for Pb²⁺ incorporating *N,N'*-bis(salicylidene)-2,6-pyridinediamine with 2-nitrophenyl octyl ether and 50 mol % lipophilic additive were reported to have rapid response and excellent selectivity towards lead

ions over other interfering metal ions [54.161]. *N,N'*-bis(5-methyl salicylidene)-*p*-diphenylene methane diamine formed a complex with Pb²⁺ and its selectivity was high, possibly due to the fact that the distance between the two nitrogen atoms and the position of the two hydroxyl group is matched with the size of Pb²⁺ ions. ISEs based upon a dispersion of this Schiff base compound into the polyvinylchloride (PVC) membrane were characterised by a fast response time, a wide linear dynamic range and a fair selectivity coefficient with a slope of 29.4 mV per decade. The sensor had a low detection limit and an active life of three months without displaying considerable divergence in potentials [54.162]. A cleaned gold microelectrode was soaked in the freshly prepared solution of *o*-amino thiophenol (*o*-AT) and the metal nitrate of Cu, Hg and Pb in *N,N*-dimethylformamide (DMF) for 4 h at room temperature. Simultaneous interaction took place between the thiophenol group of *o*-AT and the gold surface and its thiophenol and amino groups and M²⁺. The ratio of 4 : 1 was found to be appropriate for *o*-AT:M²⁺ for the self-assembly of the *o*-AT on the gold surface and the formation of recognition cavities. The prepared ISEs showed specific selectivity to the template metal ions in mixed solutions containing the three heavy-metal ions. The limits of detection for ISEs were found to be 1.46×10^{-8} M, 3.73×10^{-8} M and 4.34×10^{-8} M for Hg⁺⁺, Cu⁺⁺ and Pb⁺⁺ ions, respectively [54.163]. Membranes containing five-layer LB films of tetracarboxylic perylene derivative and polypyrrole molecules were able to detect trace levels of Cu²⁺ ions in water [54.164]. The thiol-Cu-SAM particles were doped by the adsorption of thioxyleneol or decanethiol into the porous polypyrrole film surface, resulting in an improvement in the sensitivity of the film to NH₃ [54.165]. Water-soluble regioregular polythiophenes containing acid C side chains were found to be suitable for the development of new Zn²⁺, Mn²⁺, and Cd²⁺ sensors [54.166].

54.5.3 Examples of Other Polymeric Sensors

A coordination polymer poly(CuMBSH) was formed via reaction of the bifunctional amphiphilic ligand, 5,5'-methylenebis (*N*-hexadecylsalicylideneamine) and copper ions at the water subphase. The response of LB films of poly(CuMBSH) to exposure of benzene, toluene, ethanol and water was monitored by SPR measurements. The film became swollen, possibly due to the diffusion of solvent vapour into the polymer film. Poly(CuMBSH) is regarded as a relatively nonpolar ma-

terial and its interactions were expected to be greater with vapours having a smaller dipole moment; the sensitivity increased with smaller dipole moment and higher refractive index of the vapour. The refractive index of the film increased because of an increase in the film density with the vapour filling the free volume inside the film. Similar results were obtained from admittance spectroscopic measurements on the same organic systems [54.167, 168].

The electrical properties of poly(methyl methacrylate) polymer (PMMA) films have been exploited for sensing applications (see Fig. 54.14a for the chemical formula). An increase in resistance of up to three orders of magnitude was observed for composite thin films of PMMA with carbon nanotubes when exposed to dichloromethane, chloroform and acetone. The sensing mechanism is explained on the basis of volume expansion and polar interaction of various vapours on

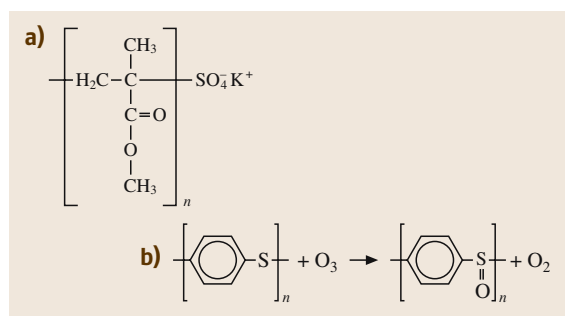


Fig. 54.14a,b Chemical structures of (a) poly(methyl methacrylate) derivatives and (b) reaction of polyphenyl-sulfide molecules with ozone

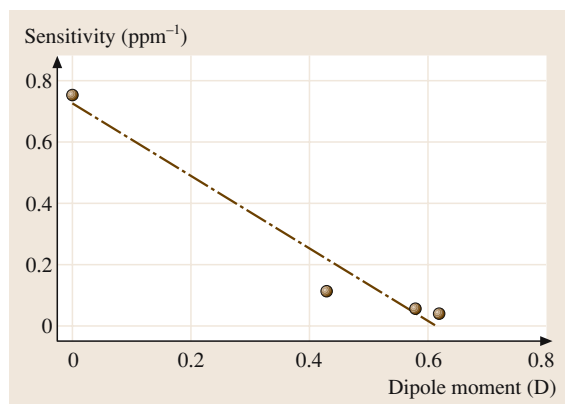


Fig. 54.15 Response of spun PMMA films on exposure to benzene (0 D), toluene (0.43 D), ethyl benzene (0.58 D) and m-xylene (0.62 D)

the nanotube surface [54.169, 170]. SPR measurements were made to investigate the room-temperature response of spun PMMA films to benzene, toluene, ethyl benzene and m-xylene (BTEX) vapours under dynamic conditions. The response of PMMA to benzene was fast and reversible. The dissolution of benzene in the bulk of the polymer film formed a heterogeneous layer, primarily giving rise to swelling of the film and the value of the partition coefficient was estimated to be 112. Figure 54.15 shows that the increase in angular shift per ppm was linear with decreasing dipole moment of the solvent vapours. Benzene is a nonpolar solvent and toluene, ethyl benzene and m-xylene are expected to show stronger solvation than benzene. Dissociated solvent ions may have formed a shell of bound ions and, as a result, the effective radius for dissolution in the polymer matrix would have increased. This dissociation might not easily occur for benzene, producing the highest SPR response [54.171]. A sensor has been made from a permselective poly(dimethylsiloxane) (PDMS) hollow-fibre membrane, showing a fast response time, very high oxygen permeability, and optimised oxygen/nitrogen selectivity. This sensor was suitable for monitoring the oxygen content in the outlet stream generated by medical oxygen concentration units [54.172].

Single-polyaniline films generally suffer from poor mechanical stability and polyaniline-blend films were prepared for practical applications using PMMA and PSt. The presence of water vapour, functioning as the interference gas, led to the lowering of the NH_3 response of the sensor in a humid atmosphere. The humidity effect was larger for PMMA than the PSt blend film. The difference was, however, not as large as expected from the water-sorption ability of PMMA [54.173]. The detection limits of ISE membranes with solid contacts are generally lower than a traditional liquid inner contact because of the diminished ion fluxes. Poly(methyl methacrylate)/poly(decyl methacrylate) (PMMA/PDMA) copolymer was employed as a solvent-cast membrane matrix on a layer of poly(*n*-octyl)thiophene (POT) deposited on Au as the internal contact. Ca^{2+} - and Pb^{2+} -selective electrodes were not sensitive to O_2 and the presence of a water film between the membrane and the internal contact was not found. The limits of detection were better by an order of magnitude and the response at low concentrations was much faster than for the liquid-contact electrode. For example, the drift was found to be smaller than 0.4 mV/min after 2 min when the concentration was increased from 10^{-9} to $10^{-7.7}$ M [54.174].

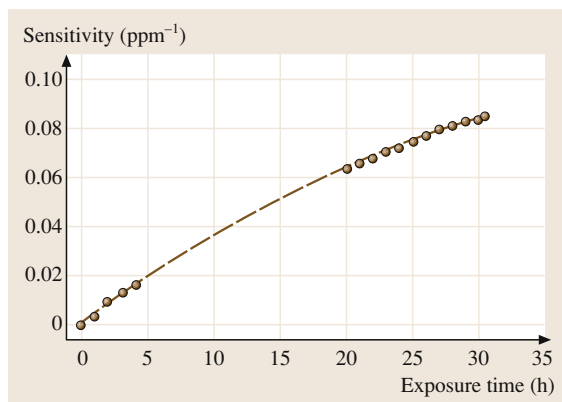


Fig. 54.16 Time-dependence behaviour of ozone sensitivity of PPS films. Two sets of hourly readings, separated by 16 h of continuous exposure without SPR measurements (dashed line)

An amperometric hydrogen sensor was proposed using a proton-conducting solid polymer electrolyte membrane (PEM), a blend of palladium and platinum as the anode and platinum as the cathode. The sensor operated as a fuel cell ($\text{H}_2/\text{Pd-Pt}/\text{PEM}/\text{Pt}/\text{O}_2$) and the short-circuit current was found to be linearly related to the hydrogen concentration [54.175].

It has recently been reported that thin films of polyphenylsulfide (PPS) (commercially termed noXon) act as efficient scrubbers of ozone from atmospheric ambient through the oxidation of the PPS molecules [54.176]. Thin films of this compound are found to be a useful material for the fabrication of

sensing membranes suitable for the detection of ground-level ozone at concentrations of a few ppm or even lower [54.177, 178]. A 1-h exposure to 2 ppm of O_3 was shown to cause at least a 1.1% change in the refractive index of the film [54.179]. As shown in Fig. 54.14b, oxidation of the PPS polymer by ozone exposure results in an irreversible increase in the index of refraction of the polymer [54.180].

Figure 54.16 shows the results of SPR measurements on PPS films when exposed to 1.5-ppm ozone. The increase in the refractive index Δn was found to be linear over the first few hours of exposure with an estimated slope of 1×10^{-4} refractive index units (RIU)/min. The sensitivity was defined as the ratio of Δn to the concentration of ozone. The interaction of O_3 with PPS films is believed to be an accumulative effect and such mechanisms can therefore serve as both a scavenger as well as an indicator of ozone in a certain ambient. The density of the binding of oxygen to a PPS film after five hours of exposure to 2-ppm ozone was estimated from QCM measurements to be about $5 \times 10^{18} \text{ cm}^{-3}$.

Electrospun nanofibers with diameters of 100–400 nm were deposited on the surface of a QCM by electrospinning homogenous blend solutions of cross-linkable poly(acrylic acid) (PAA) and poly(vinyl alcohol) (PVA). The NH_3 sensing properties were mainly affected by the content of the PAA component in the nanofibrous membranes, the concentration of NH_3 , and relative humidity. The sensitivity of nanofibrous-membrane-coated QCM sensors was also much higher than that of continuous-film-coated QCM sensors [54.181].

54.6 Cavitand Molecules

Thin films of some novel cavitand compounds such as crown ethers [54.182] and calixarenes [54.183, 184] can form inclusion complexes with some organic guest molecules. This effect is being extensively exploited for the development of sensors for organic vapours [54.185, 186]. Thin films can be prepared using LB film deposition, spin coating and self-assembly techniques [54.187–190]. The nanoporous flexible structure of thin LB films of amphiphilic calixresorcinarene derivatives are found to provide a suitable matrix for the incorporation of aromatic molecules [54.191]. This adsorption process was very fast, and full recovery of the film was observed after flushing with clean air. It has to be pointed out, however, that the detected vapours were

of a high concentration (a few percent by volume) and the adsorption was not selective since all vapours studied yielded a similar response. However, this may lead to the development of a sensor for explosives, where specificity is not considered relevant. These effects are attributed to weak and nonspecific interactions between the guest molecules and the calixarene LB film. It was also shown that the adsorption of organic vapours occurs in the whole bulk of the LB films, and that the number of adsorbed molecules is much higher than the number of calixarene molecules [54.192]. The adsorption mechanism involved the swelling of the film and even condensation of the adsorbate within the film. The swelling of the film was confirmed by independent

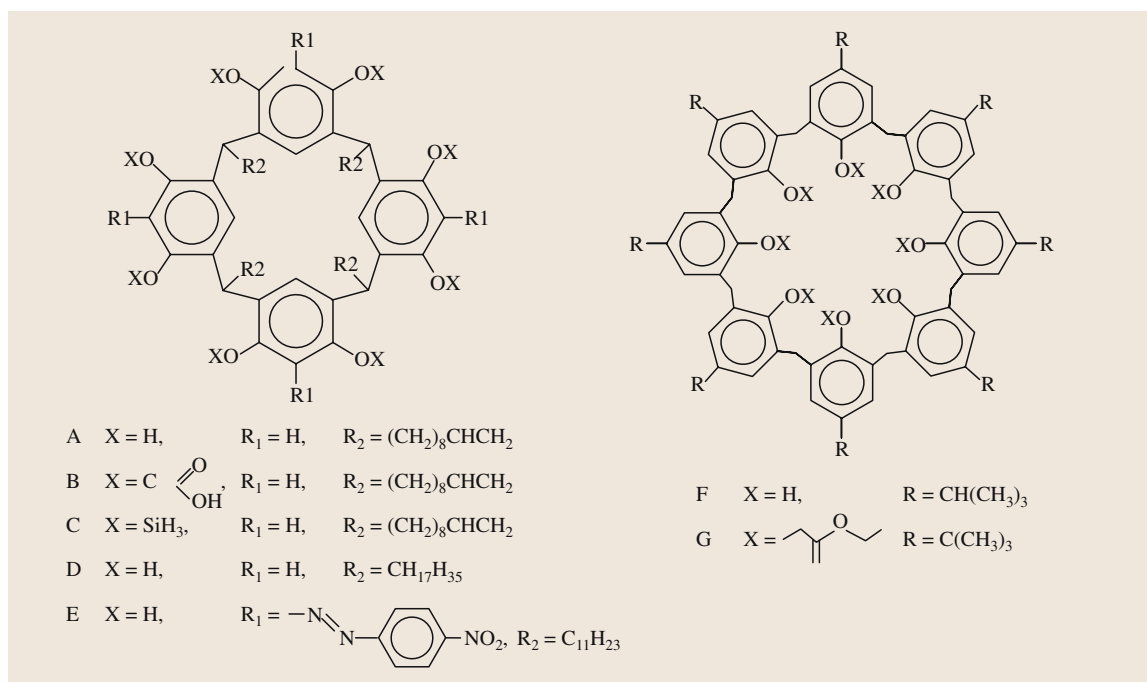


Fig. 54.17 Chemical formulae of calixresorcinarene derivatives

measurements using ellipsometric and surface plasmon resonance (SPR) techniques [54.193]; the adsorption mechanism remained still unclear. For all the studied analytes, the number of adsorbed molecules was found to be much larger than that expected from the geometrical dimensions of the intrinsic calixarene cavity and the empty space between the molecules. This discrepancy was explained by either film swelling or due to the condensation of vapour molecules inside the film, or both [54.191]. Spun films of calixresorcinarenes, however, were found to be homogeneous, and the porosity of the film was similar to that of LB-deposited films. Both LB and spun films of comparable thickness exhibited similar SPR response on exposure to toluene vapour [54.55]. Prolonged irradiation of the CA films with a focused laser beam caused an initial increase in film sensitivity to various organic vapours [54.194].

Figure 54.17 shows the chemical structure of an amphiphilic calixresorcinarene derivative with several different types of substituents. Exposures of layers of CAs to vapours of benzene, ethylbenzene, toluene and m-xylene in low concentrations produced changes in the optical properties, i. e., the film thickness and refractive index of the layers. Fast response and recovery processes have been observed, with response times as short as a few

seconds. The normalised SPR responses are shown in the three-dimensional (3-D) diagram in Fig. 54.18. It is apparent that the interaction of the analyte was, in general terms, specific to the thin film of a particular CA derivative. The adsorption mechanism was interpreted in terms of the accumulation of the vapour molecules in

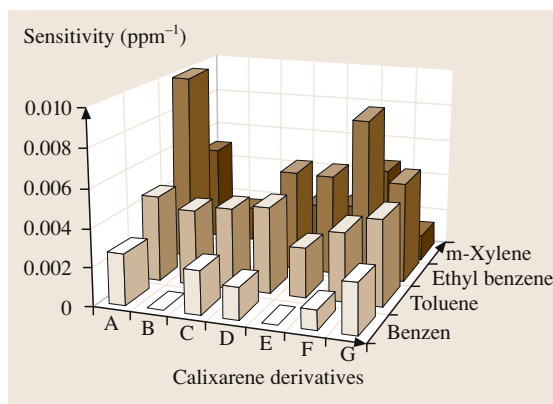


Fig. 54.18 A 3-D representation showing the response of different calixarene derivatives to different organic vapours (benzene (B), m-xylene (X), toluene (T), and ethylbenzene (E)) at a concentration of 375 ppm

the liquid phase inside the film matrix, probably caused by capillary condensation in the porous structure of the CA films [54.195]. These LB films had a characteristic pore size of about 1 nm. Condensation of vapour occurred inside capillaries similar to the porous structure of calixarene films at pressures lower than their saturated pressure at a certain temperature. The saturated vapour pressures of ethylbenzene and m-xylene are similar and therefore their exposure to the CA derivatives produced similar effects. Benzene on the other hand, with its highest saturated vapour pressure, yielded the smallest detectable SPR signal.

Composite LB films of amphiphilic CA and PPS were exposed to 2 ppm of O₃ [54.177]. The competition between the permanent oxidation process of the PPS polymer by the adsorbed O₃ molecules and the release of O₃ from the CA matrix was found to depend on the molar ratios of the CA and PPS compounds. Recovery became slower with lower CA proportions in the film matrix.

54.7 Concluding Remarks

The scope of organic materials for chemical sensing is huge and it is impossible to provide a complete description of these materials. Electronic artificial noses are being developed as systems for the automated detection and classification of odours, vapours and gases. These instruments consist of three main components: (i) an array of chemical membranes (ii) electronic circuitry for data acquisition, signal processing and display and (iii) a pattern-recognition algorithm. Up-to-date information on applications of materials for electronic noses and tongues is available in a recent review [54.197]. Problems associated with human senses are many: individual variability, impossibility of online monitoring, subjectivity, adaptation, infections, harmful exposure to hazardous compounds, and mental state. These limitations can be overcome by developing intelligent sensor systems. Gas sensors tend to have very broad selectivity, responding to many different analytes of varied concentrations. The electronic nose/tongue, therefore, offers a definite advantage in many applications in the food and utility industries, the health care and medical care sectors, and security services. The demand for advanced electronic noses will continue to stimulate the search for novel sensing materials.

Mathematical and computational tools are expected increasingly to play a part in the design of advanced

The conductivity of the CA derivative is intrinsically low. CA films were deposited on the gate between the source and the drain of a charge-flow transistor and the turn-on response for the transistor upon exposure to organic solvent was recorded. The increase in the membrane conductivity is partially attributed to condensation of the vapours in the highly microporous membrane even below the saturation vapour pressure and partially to the effect of the polar analyte molecules complexing inside and between the OH groups of the cavities. The technique offers several advantages over existing methods: (i) immunity to water vapour, due to the hydrophobic nature of the membrane; (ii) no catalytic poisoning of the membrane, as is commonly observed for doped SnO₂; (iii) no accumulative effects, which are responsible for baseline drift in quartz-crystal-type sensors, are observed, and (iv) no porous metal layer, the adhesion of which may be prone to degradation, is needed on top of the membrane [54.196].

electronic-nose systems. An electronic nose has been developed to monitor breathing air in human habitats. A molecular modelling study was undertaken to investigate the interactions between resistive sensors of a polymer-carbon black (CB) composite and analytes. Poly(4-vinylphenol), polyethylene oxide, and ethyl cellulose were considered for modelling, based on their stereoisomerism and sequence isomerism. The CB was modelled as uncharged naphthalene rings with no hydrogen. Molecular mechanical and molecular dynamics techniques were employed and the equilibrium composite structure was constructed by inserting naphthalene rings into the polymer matrix. The radial distribution profiles produced information on the composite microstructure. The sensor response was predicted in terms of the interaction energies of the analytes with the composites. Studies included both inorganic and organic analytes [54.198].

Research efforts are also growing in the mineralisation of VOC pollutants to innocuous compounds [54.199]. Visible-light-assisted removal of organic pollutants by photocatalytic action of dye-sensitised TiO₂ surfaces offers several attractive advantages. Firstly, the nanoscale device architecture is versatile with several methods of solar energy conversion whilst producing a chemical fuel in the form

of pure H₂. Secondly pollutant species and concentrations of mixtures in industrial waste streams can be identified [54.200]. Results so far reported are encouraging. For example, 40–75% degradation of aromatic pollutants such as phenol, chlorophenol, trichloroethylene and surfactants was achieved with TiO₂ surfaces modified by methylene blue and rhodamine B dyes af-

ter less than 5 h of irradiation with a 150-W xenon lamp [54.201]. Individual VOCs including methylene chloride, ethanol, benzene, acetone, xylene and isopropanol produced unique signatures as they were oxidised on the sensor surface [54.202]. Interest in hybrid hetero-supramolecular structures will therefore remain alive for the foreseeable future.

References

- 54.1 W. Gopel: *Sensors Actuat. B* **18–19**, 1–21 (1994)
- 54.2 M. C. Petty: *Biosens. Bioelectron.* **10**, 129–134 (1995)
- 54.3 M. J. Cook: *Chem. Rec.* **2**(4), 225–236 (2002)
- 54.4 L. Alcacer: *Conducting Polymers Special Applications* (Reidel Publications, Dordrecht, Holland 1987)
- 54.5 E. B. Feresenbet, E. Dalcanele, C. Dulcey, D. K. She-
noy: *Mol. Cryst. Liq. Cryst.* **397**, 585–594 (2003)
- 54.6 N. K. Hunt, B. J. Marinas: *Water Res.* **31**, 1355–1362 (1997)
- 54.7 T. Oshima, K. Sato, H. Terauchi, M. Sato: *J. Electro-
stat.* **42**, 159 (1997)
- 54.8 N. Carmona, M. A. Villegas, J. M. F. Navarro: *Sensors
Actuat. A Phys.* **116**(3), 398–404 (2004)
- 54.9 N. Carmona, M. A. Villegas, J. M. F. Navarro: *Thin
Solid Films* **458**(1–2), 121–128 (2004)
- 54.10 C. L. Baban, Y. Toyoda, M. Ogita: *Jpn. J. Appl. Phys.*
143(10), 7213–7216 (2004)
- 54.11 I. Toma-Dasu, U. Dasu, M. Karlsson: *Phys. Med.
Biol.* **49**(19), 4463–4475 (2004)
- 54.12 S. E. J. Williams, P. Wootton, H. S. Mason, J. Bould,
D. E. Iles, D. Riccardi, C. Peers, P. J. Kemp: *Science*
306(5704), 2093–2097 (2004)
- 54.13 F. Bender, C. Kim, T. Mlsna, J. F. Vetelino: *Sensors
Actuat. B Chem.* **77**(1–2), 281–286 (2001)
- 54.14 B. Onida, L. Borello, S. Fiorilli, B. Bonelli,
C. O. Arian, E. Garrone: *Chem. Commun.* **21**, 2496–
2497 (2004)
- 54.15 S. Sen, K. P. Muthe, N. Joshi, S. C. Gadkari,
S. K. Gupta, Jagannath, M. Roy, S. K. Deshpande,
J. V. Yakhmi: *Sensors Actuat. B Chem.* **98**(2–3), 154–
159 (2004)
- 54.16 B. H. Timmer, K. M. van Delft, R. P. Otjes, W. Olthuis,
A. van den Berg: *Anal. Chim. Acta* **507**(1), 137–143
(2004)
- 54.17 M. Beekmann, G. Angellet, G. Megie, H. G. J. Smith,
D. Kley: *J. Atmos. Chem.* **19**, 259 (1994)
- 54.18 L. Xie, T. J. Lu, H. Q. Yan: *Electroanal.* **10**, 842 (1998)
- 54.19 Th. Becker, L. Tomasi, Chr. Bosch-v.Braunmühl,
G. Müller, G. Sberveglieri, G. Fagli, E. Comini: *Sen-
sors Actuat. A Phys* **74**, 229 (1999)
- 54.20 H. Schulz, G. B. De Melo, F. Ousmanov: *Combust.
Flame* **118**, 179–190 (1999)
- 54.21 B. Zielinska, J. C. Sagebiel, G. Harshfield, A. W. Gert-
ler, W. R. Pierson: *Atmos. Environ.* **30**, 2269–86
(1996)
- 54.22 F. B. Reig, J. V. Adelantado, V. P. Martinez,
M. C. Moreno, M. T. Carbo: *J. Molec. Struct.* **480–481**,
529–534 (1999)
- 54.23 W. Groves, E. T. Zellers, G. C. Frye: *Anal. Chim. Acta*
371, 131 (1998)
- 54.24 W. A. Groves, E. T. Zellers: *Ann. Occup. Hyg.* **45**,
609–623 (2001)
- 54.25 N. Kasai, I. Sugimoto, M. Nakamuro, T. Katoh:
Biosens. Bioelectron. **14**, 533–539 (1999)
- 54.26 G. Sberveglieri: *Sensors Actuat. B* **23**, 103–109 (1995)
- 54.27 D. Manno, A. Serra, M. Di Giulio, G. Micocci,
A. Tepore: *Thin Solid Films* **324**, 44–51 (1998)
- 54.28 N. Taguchi: Japanese Patent Application No. 45–
38200 (1962)
- 54.29 C. D. Kohl, A. Eberheim, P. Schieberle: *Tech. Mess.*
71(5), 298–304 (2004)
- 54.30 A. Eberheim, D. Kohl, P. Schieberle: *Phys. Chem.
Chem. Phys.* **5**(23), 5203–5206 (2003)
- 54.31 J. Gutierrez, J. Getino, M. C. Horrillo, L. Ares,
J. I. Robla, C. Garcia, I. Sayago: *Thin Solid Films* **317**,
429–431 (1998)
- 54.32 M. Graf, D. Barretino, S. Taschini, C. Hagleitner,
A. Hierlemann, H. Baltés: *Anal. Chem.* **76**(15), 4437–
4445 (2004)
- 54.33 E. Comini, G. Faglia, G. Sberveglieri, Zhengwei Pan,
Zhong L. Wang: *Appl. Phys. Letts.* **81**(10), 1869–1871
(2002)
- 54.34 A. Gramm, A. Schutze: *Sensors Actuat. B Chem.*
95(1–3), 58–65 (2003)
- 54.35 O. K. Varghese, D. Gong, M. Paulose, K. G. Ong,
C. A. Grimes: *Sensors Actuat. B Chem.* **93**, 338–344
(2003)
- 54.36 A. K. Prasad, P. I. Gouma: *J. Mater. Sci.* **38**(21), 4347–
4352 (2003)
- 54.37 N. B. McKeown: In: *Phthalocyanine Materials. Syn-
thesis, Structure and Function* (Cambridge Univ.
Press, Cambridge 1998) p. 60
- 54.38 S. Tabuchi, H. Tabata, T. Kawai: *Surf. Sci.* **571**(1–3),
117–127 (2004)
- 54.39 M. Szybowski, T. Runka, M. Drozdowski, W. Bala,
A. Grodzicki, P. Piszczek, A. Bratkowski: *J. Molec.
Struct.* **704**(1–3), 107–113 (2004)
- 54.40 A. Boguta, D. Wrobel, A. Bartczak, R. Swietlik,
Z. Stachowiak, I. M. Ion: *Mater. Sci. Eng. B Solid*
113(1), 99–105 (2004)

- 54.41 J. Spadavecchia, G. Ciccarella, S. Capone, R. Rella: Chem. Mater. **16**(11), 2083–2090 (2004)
- 54.42 H. Nishimura, M. Iizuka, S. Kuniyoshi, M. Nakamura, K. Kudo, K. Tanaka: Electron. Commun. Jpn. **87**(2), 18–25 (2004)
- 54.43 Y. L. Lee, H. Y. Wu, C. H. Chang, Y. M. Yang: Thin Solid Films **423**(2), 169–177 (2003)
- 54.44 M. C. Petty: *Langmuir–Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge 1996)
- 54.45 M. J. Cook: J. Mater. Sci. Electron. **5**, 117–128 (1994)
- 54.46 M. J. Cook: Int. J. Electron. **76**, 727–739 (1994)
- 54.47 C. G. Claessens, W. J. Blau, M. Cook, M. Hanack, R. J. M. Nolte, T. Torres, D. Wohrle: Monatsh. Chem. **132**(1), 3–11 (2001)
- 54.48 M. J. Cook: J. Mater. Chem. **6**, 677–689 (1996)
- 54.49 T. H. Richardson, C. M. Dooling, O. Worsfold, L. T. Jones, K. Kato, K. Shinbo, F. Kaneko, R. Treginning, M. O. Vysotsky, C. A. Hunter: Colloid Surf. A **198**, 843–857 (2002)
- 54.50 M. J. Cook, J. McMurdo, D. A. Miles, R. H. Poynter, J. M. Simmons, S. D. Haslam, R. M. Richardson, K. Welford: J. Mater. Chem. **4**, 1205–1213 (1994)
- 54.51 A. K. Hassan, A. K. Ray, A. V. Nabok, S. Panigrahi: IEE Proc.–Sci. Meas. Technol. **147**, 137–140 (2000)
- 54.52 D. Meyerhofer: J. Appl. Phys. **49**, 3993–7 (1978)
- 54.53 P. C. Sukaneck: J. Electrochem. Soc. **138**, 1712–1719 (1991)
- 54.54 P. C. Sukaneck: J. Electrochem. Soc. **144**, 3959–3962 (1997)
- 54.55 A. K. Hassan, A. V. Nabok, A. K. Ray, A. Lucke, K. Smith, C. J. M. Stirling, F. Davis: Supramol. Sci. Mater. Sci. Eng. C **8–9**, 251–255 (1999)
- 54.56 S. M. Critchley, M. R. Willis, M. J. Cook, J. McMurdo, Y. Maruyama: J. Mater. Chem. **2**, 157 (1992)
- 54.57 G. C. Bryant, M. J. Cook, C. Ruggier, T. G. Ryan, A. J. Thorne, S. D. Haslam, R. M. Richardson: Thin Solid Films **243**, 316–324 (1994)
- 54.58 X. Li, H. Xu, Q. Zhou, D. Jiang, L. Zhang, A. Lu: Thin Solid Films **324**, 277–280 (1998)
- 54.59 K. Bandyopadhyay, S. G. Liu, H. Y. Liu, L. Echegoyen: Chem. Eur. J. **6**, 4385–4392 (2000)
- 54.60 C. C. Hsueh, M. T. Lee, M. S. Freund, G. S. Ferguson: Angew. Chem. Int. Ed. **39**, 1228–1230 (2000)
- 54.61 H. G. Hong, M. Jiang, S. G. Sligar, P. W. Bohn: Langmuir **10**, 153–8 (1994)
- 54.62 K. F. Kelly, Y. B. S. Shon, T. R. Lee, N. J. Halas: J. Phys. Chem. B **103**, 8639 (1999)
- 54.63 V. Poderys, A. Selskis, R. Rotomskis: Solid State Phenom. **97–98**, 221–224 (2004)
- 54.64 A. Ulman: *An Introduction to Ultrathin Films: From Langmuir–Blodgett to Self-Assembly* (Academic, San Diego 1991)
- 54.65 T. R. E. Simpson, D. A. Russell, I. Chambrier, M. J. Cook, A. B. Horn, S. C. Thorpe: Sensors Actuat. B **29**, 353–357 (1995)
- 54.66 M. J. Cook: Pure Appl. Chem. **71**(11), 2145–2151 (1999)
- 54.67 K. Ozoemena, P. Westbroek, T. Nyokong: J. Porphyr. Phthalocyan. **6**(2), 98–106 (2002)
- 54.68 X. Zhang, J. C. Shen: Adv. Mater. **11**(13), 1139–1143 (1999)
- 54.69 Y. Lvov, G. Decher, H. Möhwald: Langmuir **9**, 481 (1993)
- 54.70 Y. Lvov, K. Ariga, I. Ichinose, T. Kunitake: J. Am. Chem. Soc. **117**, 6117 (1995)
- 54.71 Y. M. Lvov, G. Decher: Crystallogr. Rep. **39**, 628 (1994)
- 54.72 X. Zhang, M. L. Gao, X. X. Kong, Y. P. Sun, J. C. Shen: J. Chem. Soc. Chem. Commun. **9**, 1055–1056 (1994)
- 54.73 C. Q. Sun, X. Y. Zhang, D. Jiang, Q. A. Gao, H. D. Xu, Y. P. Sun, X. Zhang, J. C. Shen: J. Electroanal. Chem. **411**(1–2), 73–78 (1996)
- 54.74 D. J. Revell, I. Chambrier, M. J. Cook, D. A. Russell: J. Mater. Chem. **10**(1), 31–37 (2000)
- 54.75 M. J. Cook, I. Chambrier: Phthalocyanine properties. In: *Porphyrin Handbook*, Vol. 15, ed. by K. Kadish et al. (Academic, New York 2003) Chap. 108
- 54.76 S. Antohe, N. Tomozeiu, S. Gogonea: Phys. Stat. Sol. (a) **125**, 397–408 (1991)
- 54.77 M. J. Cook, M. F. Daniel, K. J. Harrison, N. B. Mckee, A. J. Thomson: J. Chem. Soc. Chem. Commun. **15**, 1148–1150 (1987)
- 54.78 M. J. Cook, M. F. Daniel, K. J. Harrison, N. B. Mckee, A. J. Thomson: J. Chem. Soc. Chem. Commun. **14**, 1086–1088 (1987)
- 54.79 A. K. Ray, A. V. Nabok, A. K. Hassan, O. Omar, R. Taylor, M. J. Cook: Philos. Mag. B **78**(1), 53–64 (1998)
- 54.80 X. B. Huang, Y. Q. Liu, S. Wang, S. Q. Zhou, D. B. Zhu: Chem. Eur. J. **8**(18), 4179–4184 (2002)
- 54.81 C. Granito, L. M. Goldenberg, M. R. Bryce, A. P. Monkman, L. Troisi, L. Pasimeni, M. C. Petty: Langmuir **12**(2), 472–476 (1996)
- 54.82 B. N. Achar, P. K. Jayasree: Can. J. Chem./Rev. Can. Chim. **77**(10), 1690–1696 (1999)
- 54.83 H. R. Kerp, E. E. van Faassen: Proceedings of the 11th Workshop on Quantum Solar Energy Conversion – (QUANTSOL'98), Chem. Phys. Lett. **332**, 5 (2000)
- 54.84 P. D. Hooper, M. I. Newton, G. McHaleand, M. R. Willis: Semicond. Sci. Technol. **12**, 455–459 (1997)
- 54.85 D. Markovitsi, T. H. Tran-Thi, R. Even, J. Simon: Chem. Phys. Lett. **137**, 107 (1987)
- 54.86 T. Basova, E. Kol'tsov, A. K. Hassan, A. Nabok, A. G. Gurek. Ray, V. Ahsen: J. Mater. Sci.–Mater. El. **15**(9), 623–628 (2004)
- 54.87 W. Snow, W. R. Barger: Phthalocyanine films in chemical sensors. In: *Phthalocyanines. Properties and Applications*, ed. by C. C. Leznoff, A. B. P. Lever (VCH, New York 1989)
- 54.88 M. Nicolau, B. del Rey, T. Torres, C. Mingotaud, P. Delhaes, M. J. Cook, S. C. Thorpe: Synth. Met. **102**(1–3), 1462–1463 (1999)
- 54.89 R. Zhou, F. Josse, W. Gopel, Z. Z. Öztürk, Ö. Bekaroğlu: Appl. Organomet. Chem. **10**, 557–577 (1996)

- 54.90 A. Chyla, A. Lewandowska, J. Soloducho, A. Gorecka-Drzazga, M. Szablewski: IEEE Trans. Dielect. El In **8**(3), 559–565 (2001)
- 54.91 S. Gao, H. Zhao, L. H. Huo, J. G. Zhao, Y. Q. Wu, S. Q. Xi: Sensors Actuat. B Chem. **97**(2–3), 319–323 (2004)
- 54.92 S. Suslick, N. A. Rakow, M. E. Kosal, J.-H. Chou: J. Porphyr. Phthalocyan. **4**, 407–413 (2001)
- 54.93 Y. Lee, B. K. Oh, M. E. Meyerhoff: Anal. Chem. **76**(3), 536–544 (2004)
- 54.94 J. Charvatova, O. Rusin, V. Kral, K. Volka, P. Matejka: Sensors Actuat. B **76**(1–3), 366–372 (2001)
- 54.95 O. Ikeda, H. Koyama, K. Kijima, T. Komura, A. Itajima, M. Miyake, K. Yamamoto, A. Yamatodani: *Proceedings of the 27th Chemical Sensor Symposium*, Vol. 14 (Supplement B) (Japan Association of Chemical Sensor, Nagaoka University of Technology October 23–24, 1998) pp. 89–92
- 54.96 V. C. Smith, T. Richardson, H. L. Anderson: Supramolec. Sci. **4**(3–4), 503–508 (1997)
- 54.97 V. Arima, R. I. R. Blyth, F. Della Sala, R. Del Sole, F. Matino, G. Mele, G. Vasapollo, R. Cingolani: Mater. Sci. Eng. C Bio. Solids **24**(4), 569–573 (2004)
- 54.98 H. Imahori, K. Hosomizu, Y. Mori, T. Sato, T. K. Ahn, S. K. Kim, D. Kim, Y. Nishimura, I. Yamazaki, H. Ishii, H. Hotta, Y. Matano: J. Phys. Chem. B **108**(16), 5018–5025 (2004)
- 54.99 Q. L. Li, S. Surthi, G. Mathur, S. Gowda, Q. Zhao, T. A. Sorenson, R. C. Tenent, K. Muthukumar, J. S. Lindsey, V. Misra: Appl. Phys. Lett. **85**(10), 1829–1831 (2004)
- 54.100 K. E. Splan, J. T. Hupp: Langmuir **20**(24), 10560–10566 (2004)
- 54.101 J. R. C. da Rocha, G. J. F. Demets, M. Bertotti, K. Araki, H. E. Toma: J. Electroanal. Chem. **526**(1–2), 69–76 (2002)
- 54.102 J. D. Wright: Prog. Surf. Sci. **31**(1–2), 1–60 (1989)
- 54.103 J. P. Germain, A. Pauly, C. Maleysson, J. P. Blanc, B. Schöllhorn: Thin Solid Films **333**, 235–239 (1998)
- 54.104 B. Schöllhorn, J. P. Germain, A. Pauly, C. Maleysson, J. P. Blanc: Thin Solid Films **326**, 245–250 (1998)
- 54.105 J. Travis, A. K. Ray, S. C. Thorpe, M. J. Cook, S. A. James: Meas. Sci. Technol. **6**(7), 988–994 (1995)
- 54.106 A. Cole, R. J. McIlroy, S. C. Thorpe, M. J. Cook, J. McMurdo, A. K. Ray: Sensors Actuat. B **13**, 416–419 (1993)
- 54.107 D. Crouch, S. C. Thorpe, M. J. Cook, I. Chambrier, A. K. Ray: Sensors Actuat. B **18–19**, 411–414 (1994)
- 54.108 A. Tepore, A. Serra, D. P. Arnold, D. Manno, G. Micocci, A. Genga, L. Valli: Langmuir **17**(26), 8139–8144 (2001)
- 54.109 Y. L. Lee, C. Y. Sheu, R. H. Hsiao: Sensors Actuat. B Chem. **99**(2–3), 281–287 (2004)
- 54.110 Y. L. Lee, C. H. Hsiao, C. H. Chang, Y. M. Yang: Sens. Actuat. B **94**, 169–175 (2003)
- 54.111 M. I. Newton, T. K. H. Starke, G. McHale, M. R. Willis: Thin Solid Films **360**(1–2), 10–12 (2000)
- 54.112 M. I. Newton, T. K. H. Starke, M. R. Willis, G. McHale: Sens. Actuat. B **67**, 307–311 (2000)
- 54.113 Q. Zhou, R. D. Gould: Thin Solid Films **317**(1–2), 436–439 (1998)
- 54.114 W. F. Qiu, W. P. Hu, Y. Q. Liu, S. Q. Zhou, Y. Xu, D. B. Zhu: Sensors Actuat. B Chem. **75**(1–2), 62–66 (2001)
- 54.115 M. Bouvet, A. Leroy, J. Simon, F. Tournilhac, G. Guillaud, P. Lessnick, A. Maillard, S. Spirkovitch, M. Debligny, A. Haan, A. Decroly: Sensors Actuat. B Chem. **72**(1), 86–93 (2001)
- 54.116 M. Bouvet, G. Guillaud, A. Leroy, A. Maillard, S. Spirkovitch, F. G. Tournilhac: Sensors Actuat. B Chem. **73**(1), 63–70 (2001)
- 54.117 C. Q. Sun, Y. P. Sun, X. Zhang, H. D. Xu, J. C. K. Shen: Anal. Chim. Acta **312**(2), 207–212 (1995)
- 54.118 Y. P. Sun, X. Zhang, C. Q. Sun, Z. Q. Wang, J. C. Shen, D. J. Wang, T. J. Li: Chem. Commun. **20**, 2379–2380 (1996)
- 54.119 K. F. Schoch, J. Gregg, T. A. Temofonte: J. Vac. Sci. Technol. A **6**(1), 155–158 (1988)
- 54.120 A. K. Hassan, A. K. Ray, J. R. Travis, Z. Ghassemloooy, M. J. Cook, A. Abass, R. A. Collins: Sensors Actuat. B Chem. **49**(3), 235–239 (1998)
- 54.121 J. Mårtensson, H. Arwin, I. Lundstrom: Sensors Actuat. B Chem. **1**(1–6), 134–137 (1990)
- 54.122 J. M. Pedrosa, C. M. Dooling, T. H. Richardson, R. K. Hyde, C. A. Hunter, M. T. Martin, L. Camacho: J. Mater. Chem. **12**(9), 2659–2664 (2002)
- 54.123 O. Worsfold, C. M. Dooling, T. H. Richardson, M. O. Vysotsky, R. Tregonning, C. A. Hunter, C. Mallins: Colloid Surf. A **198**, 859–867 (2002)
- 54.124 L. Gaffo, O. D. D. Couto, R. Giro, M. J. S. P. Brasil, D. S. Galvao, F. Cerdeira, O. N. de Oliveira, K. Wohnerath: Solid State Commun. **131**(1), 53–56 (2004)
- 54.125 T. Richardson, V. C. Smith, A. Topaci, J. Jiang, C. H. Huang: Supramol. Sci. **4**, 465–470 (1997)
- 54.126 T. R. E. Simpson, M. J. Cook, M. C. Petty, S. C. Thorpe, D. A. Russell: Analyst **121**(10), 1501–1505 (1996)
- 54.127 T. R. E. Simpson, D. J. Revell, M. J. Cook, D. A. Russell: Langmuir **13**(3), 460–464 (1997)
- 54.128 E. Kretschmann: Z. Phys. **241**, 313–324 (1971)
- 54.129 I. Pockrand: Surf. Sci. **72**, 577–588 (1978)
- 54.130 J. M. Rooney, E. A. H. Hall: Anal. Chem. **76**(23), 6861–6870 (2004)
- 54.131 T. Basova, E. Kol'tsov, A. Hassan, A. Tsargorodskaya, A. K. Ray, I. Igumenov: Phys. Stat. Sol. (b) **242**(4), 822 (2005)
- 54.132 J. D. Wright, A. Cado, S. J. Peacock, V. Rivalle, A. M. Smith: Sens. Actuat. B **29**, 108–114 (1995)
- 54.133 J. P. Lloyd, C. Pearson, M. C. Petty: Thin Solid Films **160**(1–2), 431–443 (1988)
- 54.134 M. J. Jory, P. S. Cann, J. R. Sambles: J. Phys. D Appl. Phys. **27**(1), 169–174 (1994)
- 54.135 K. Kato, C. M. Dooling, K. Shinbo, T. H. Richardson, F. Kaneko, R. Tregonning, M. O. Vysotsky, C. A. Hunter: Colloid Surf. A **198**, 811–816 (2002)

- 54.136 D. S. Ballantine, R. M. White, S. I. Martin, A. J. Ricco, E. T. Zellers, G. C. Fry, H. Wohltjen: *Acoustic Wave Sensors. Theory, Design, and Physico-Chemical Applications* (Academic, New York 1997)
- 54.137 K. D. Schierbaum, R. Zhou, S. Knecht, R. Dieing, M. Hanack, W. Göpel: *Sensors Actuat. B Chem.* **24**, 69–71 (1995)
- 54.138 H. Ding, V. Erokhin, M. K. Ram, S. Paddeu, L. Valkova, C. Nikolini: *Thin Solid Films* **379**, 279–286 (2000)
- 54.139 Z. Z. Öztürk, R. Zhou, U. Weimar, V. Ahsen, O. Bekaroğlu, W. Göpel: *Sensors Actuat. B Chem.* **26–27**, 208–212 (1995)
- 54.140 J. Souto, M. L. Rodriguez, J. A. Desaja, R. Aroca: *Int. J. Electron.* **76**(5), 763–769 (1994)
- 54.141 T. Basova, C. Tasaltin, A. G. Gurek, M. A. Ebeoğlu, Z. Z. Öztürk, V. Ahsen: *Sensors Actuat. B Chem.* **96**(1–2), 70–75 (2003)
- 54.142 C. Barriain, I. R. Matias, C. Fernandez-Valdivielso, F. J. Arregui, M. L. Rodriguez-Mendez, J. A. de Saja: *Sensors Actuat. B Chem.* **93**(1–3), 153–158 (2003)
- 54.143 J. Spadavecchia, G. Ciccarella, R. Rella, S. Capone, P. Siciliano: *Sensors Actuat. B Chem.* **96**(3), 489–497 (2003)
- 54.144 J. Spadavecchia, G. Ciccarella, A. Buccolieri, G. Vasapollo, R. Rella: *J. Porphy. Phthalocyan.* **7**(8), 572–578 (2003)
- 54.145 S. Nardis, D. Monti, C. Di Natale, A. D'Amico, P. Siciliano, A. Forleo, M. Epifani, A. Taurino, R. Rella, R. Paollesse: *Sensors Actuat. B Chem.* **103**(1–2), 339–343 (2004)
- 54.146 J. Spadavecchia, G. Ciccarella, G. Vasapollo, P. Siciliano, R. Rella: *Sensors Actuat. B Chem.* **100**(1–2), 135–138 (2004)
- 54.147 C. Granito, J. N. Wilde, S. Houghton, P. J. Iredale: *Thin Solid Films* **284–285**, 98–101 (1996)
- 54.148 G. Jin, C. O. Too, J. Norrish, G. G. Wallace: *Synth. Met.* **135**(1–3), 29–30 (2003)
- 54.149 V. Papes, S. Brodska: *Sensors Actuat. B Chem.* **40**, 143–145 (1997)
- 54.150 A. Riul, A. M. G. Soto, S. V. Mello, S. Bone, D. M. Taylor, L. H. C. Mattoso: *Synth. Met.* **132**(2), 109–116 (2003)
- 54.151 J. S. Do, W. B. Chang: *Sensors Actuat. B Chem.* **101**(1–2), 97–106 (2004)
- 54.152 G. Anitha, E. Subramanian: *Sensors Actuat. B Chem.* **92**(1–2), 49–59 (2003)
- 54.153 E. Segal, R. Tchoudakov, M. Narkis, A. Siegmann, Y. Wei: *Sensors Actuat. B Chem.* **104**(1), 140–150 (2005)
- 54.154 E. Milella, F. Musio, M. B. Alba: *Thin Solid Films* **285**, 908 (1996)
- 54.155 Y. Kunugi, K. Nigorikawa, Y. Harima, K. Yamashita: *J. Chem. Soc. Chem. Commun.* **Issue 7**, 873 (1994)
- 54.156 S. Christie, E. Scorsone, K. Persaud, F. Kvasnik: *Sensors Actuat. B Chem.* **90**(1–3), 163–169 (2003)
- 54.157 M. Ando, C. Swart, E. Pringsheim, V. M. Mirsky, O. S. Wolfbeis: *Solid State Ionics* **152–153**, 819–822 (2002)
- 54.158 A. Penza, G. Cassano, A. Sergi, C. Lo Sterzo, M. Russo: *Sensors Actuat. B Chem.* **81**(1), 88–98 (2001)
- 54.159 T. A. Dickinson, J. White, J. S. Kauer, D. R. Walt: *Nature* **382**, 697 (1996)
- 54.160 D. Li, C. A. Mills, J. M. Cooper: *Sensors Actuat. B Chem.* **92**(1–2), 73–80 (2003)
- 54.161 T. Jeong, H. K. Lee, D. C. Jeong, S. Jeon: *Talanta* **65**(2), 543–548 (2005)
- 54.162 M. M. Ardakany, A. A. Ensafi, H. Naeimi, A. Dastanpour, A. Shamli: *Sensors Actuat. B Chem.* **96**(1–2), 441–445 (2003)
- 54.163 S. Y. Huan, C. X. Jiao, Q. Shen, J. H. Jiang, G. M. Zeng, H. H. Guo, S. L. Guo, R. Q. Yu: *Electrochim. Acta* **49**(25), 4273–4280 (2004)
- 54.164 P. A. Antunes, C. M. Santana, R. F. Aroca, O. N. Oliveira, C. J. L. Constantino, A. Riul: *Synth. Met.* **148**(1), 21–24 (2005)
- 54.165 T. Shimanouchi, S. Morita, H. S. Jung, Y. Sakurai, Y. Suzuki, R. Kuboi: *Sensor Mater.* **16**(5), 255–265 (2004)
- 54.166 P. C. Ewbank, R. S. Loewe, L. Zhai, J. Reddinger, G. Sauve, R. D. McCullough: *Tetrahedron* **60**(49), 11269–11275 (2004)
- 54.167 J. N. Wilde, J. Nagel, M. C. Petty: *Thin Solid Films* **327–329**, 726–729 (1998)
- 54.168 R. Casalini, J. N. Wilde, J. Nagel, U. Oertel, M. C. Petty: *Sensors Actuat. B* **57**, 28–34 (1999)
- 54.169 J. K. Abraham, B. Philip, A. Witchurch, V. K. Varadan, C. C. Reddy: *Smart Mater. Struct.* **13**(5), 1045–1049 (2004)
- 54.170 B. Philip, J. K. Abraham, A. Chandrasekhar, V. K. Varadan: *Smart Mater. Struct.* **12**(6), 935–939 (2003)
- 54.171 R. Capan, A. K. Ray, A. K. Hassan, T. Tanrisever: *J. Phys. D Appl. Phys.* **36**, 1115–1119 (2003)
- 54.172 R. Rego, N. Caetanoc, R. Vale, A. Mendes: *J. Membr. Sci.* **244**(1–2), 35–44 (2004)
- 54.173 M. Matsuguchi, A. Okamoto, Y. Sakai: *Sensors Actuat. B Chem.* **94**(1), 46–52 (2003)
- 54.174 J. Sutter, A. Radu, S. Peper, E. Bakker, E. Pretsch: *Anal. Chim. Acta* **523**(1), 53–59 (2004)
- 54.175 C. Ramesh, G. Velayutham, N. Murugesan, V. Ganesan, V. Manivannan, G. Periaswami: *Ionics* **10**(1–2), 50–55 (2004)
- 54.176 A. Calogirou, M. Duan, D. Kotzias, M. Lahaniati, B. R. Larsen: *Atmos. Environ.* **31**, 2741 (1997)
- 54.177 A. V. Nabok, A. K. Hassan, A. K. Ray, J. Travis, M. Hofton, A. Dalley: *IEE Proc. Sci., Meas. Technol.* **147**, 153 (2000)
- 54.178 A. K. Ray, A. V. Nabok, A. K. Hassan, M. Hofton, A. Dalley: *Sensors & Their Applications X Conference*, ed. by N. M. White, J. T. Augousti 1999

- 54.179 A. K. Hassan, A. V. Nabok, A. K. Ray, G. Kiouisis: Mater. Sci. Eng. C **22**, 197–200 (2002)
- 54.180 A. K. Hassan, J. Greenway, A. K. Ray, A. V. Nabok: J. Phys. D Appl. Phys. **36**(17), 2130–2133 (2003)
- 54.181 B. Ding, J. H. Kim, Y. Miyazaki, S. M. Shiratori: Sensors Actuat. B Chem. **101**(3), 373–380 (2004)
- 54.182 Z. Cao, D. Cao, Z. G. Lei, H. G. Lin, R. Q. Yu: Talanta **44**, 1413 (1997)
- 54.183 C. D. Gutsche: *Calixarenes* (Royal Society of Chemistry, Cambridge 1989)
- 54.184 D. J. Cram, S. Karbach, H.-E. Kim, C. B. Knobler, E. F. Maverick, J. L. Ericson, R. C. Hegelson: J. Am. Chem. Soc. **110**, 2229 (1988)
- 54.185 F. L. Dickert, U. P. A. Baumler, G. K. Zwissler: Synth. Met. **61**, 47 (1993)
- 54.186 P. Nelli, E. Delcanale, G. Faglia, G. Sberveglieri, P. Soncini: Sensors Actuat. B Chem. **13–14**, 302 (1993)
- 54.187 T. Weiss, K. D. Schierbaum, W. Göpel, U. Thoden van Velzen, D. N. Reinholdt: Sensors Actuat. B Chem. **26**, 203 (1995)
- 54.188 E. Dalcanale, J. Hartman: Sensors Actuat. B Chem. **24**, 39 (1995)
- 54.189 J. Rickert, T. Weiss, W. Göpel: Sensors Actuat. B Chem. **31**, 45 (1996)
- 54.190 A. K. Hassan, A. K. Ray, A. V. Nabok, F. Davis: Sensors Actuat. B **77**, 638–641 (2001)
- 54.191 A. V. Nabok, N. V. Lavrik, Z. I. Kazantseva, B. A. Nesterenko, L. N. Markovskiy, V. I. Kalchenko, A. V. Shvianiuk: Thin Solid Films **259**, 244–247 (1995)
- 54.192 A. V. Nabok, A. K. Hassan, A. K. Ray: J. Mater. Chem. **10**, 189–194 (2000)
- 54.193 A. V. Nabok, A. K. Hassan, A. K. Ray, O. Omar, V. I. Kalchenko: Sensors Actuat. B Chem. **45**, 115 (1997)
- 54.194 A. K. Hassan, A. K. Ray, A. V. Nabok, T. Wilkop: Appl. Surf. Sci. **182**, 49–54 (2001)
- 54.195 S. J. Gregg, K. S. W. Sing: *Adsorption, Surface Area and Porosity* (Academic, New York 1967)
- 54.196 T. Wilkop, A. K. Ray: J. Phys. D Appl. Phys. **35**(20), 2661–2667 (2002)
- 54.197 P. Gouma, G. Sberveglieri, R. Dutta, J. W. Gardner, E. L. Hines: MRS Bull. **29**(10), 697–700 (2004)
- 54.198 A. V. Shevade, M. A. Ryan, M. L. Homer, A. M. Manfreda, H. Zhou, K. S. Manatt: Sensors Actuat. B Chem. **93**(1–3), 84–91 (2003)
- 54.199 A. Mills, S. L. Hunte: J. Photochem. Photobiol. A **108**, 1–35 (1997)
- 54.200 G. J. Wilson, G. D. Will: Curr. Appl. Phys. **4** (2–4), 351–354 (2004)
- 54.201 D. Chatterjee, A. Mahata: J. Photochem. Photobiol. **153**(1–3), 199–204 (2002)
- 54.202 L. R. Skubal, N. K. Meshkov, M. C. Vogt: J. Phototechnol. Photobiol. A **148**(1–3), 103–108 (2002)

Packaging Materials

This chapter is a high-level overview of the materials used in an electronic package including: metals used as conductors in the package, ceramics and glasses used as dielectrics or insulators and polymers used as insulators and, in a composite form, as conductors. There is a need for new materials to meet the ever-changing requirements for high-speed digital and radio-frequency (RF) applications. There are different requirements for digital and RF packages that translate into the need for unique materials for each application. The interconnect and dielectric (insulating) requirements are presented for each application and the relevant materials properties and characteristics are discussed. The fundamental materials characteristics are: dielectric constant, dielectric loss, thermal and electric conductivity, resistivity, moisture absorption, glass-transition temperature, strength, time-dependent deformation (creep), and fracture toughness. The materials characteristics and properties are dependant on how they are processed to form the electronic package so the fundamentals of electronic packaging processes are discussed including wirebonding, solder interconnects, flip-chip interconnects, underfill for flip chip and overmolding. The relevant materials properties are given along with requirements (including environmentally friendly Pb-free packages) that require new materials to be developed to meet future electronics needs for both digital and RF applications.

55.1 Package Applications	1268
55.2 The Materials Challenge of Electronic Packaging	1269
55.2.1 Materials Issues in High-Speed Digital Packaging ..	1270
55.2.2 RF Packaging Materials Issues	1271
55.3 Materials Coefficient of Thermal Expansion	1272
55.4 Wirebond Materials	1272
55.4.1 Wirebonds for Digital Applications	1272
55.4.2 Wirebonds for RF	1273
55.5 Solder Interconnects	1273
55.5.1 Flip-Chip Interconnects	1276
55.5.2 Flip Chip for RF	1277
55.5.3 Pb-Free	1277
55.6 Substrates	1278
55.6.1 RF Substrate Materials	1279
55.7 Underfill and Encapsulants	1280
55.7.1 Underfill	1280
55.7.2 Encapsulation	1280
55.8 Electrically Conductive Adhesives (ECAs) ..	1281
55.8.1 Adhesive Polymers	1281
55.8.2 Metal Fillers	1282
55.8.3 Conduction Mechanisms	1282
55.8.4 Isotropic Versus Anisotropic Conduction	1282
55.8.5 Rework	1283
55.9 Thermal Issues	1283
55.9.1 Thermal Issues in Digital Packaging	1283
55.9.2 Thermal Issues in RF Packaging ...	1284
55.10 Summary	1284
References	1285

An electronic package is a configuration of materials that interconnects electronic signals from one area to another. This scheme must isolate these signals so that there is no interference and must also protect the electronics from a degrading environment. A package typically consists

of the semiconductor, mounted and interconnected (with solder or Au wires) to a dielectric substrate (with a lead-frame or with metal traces), which is encapsulated to seal the device from the environment. The electronic package must serve four functions:

- provide electrical (or photonic) contact to and from the chip,
- act as an electrical space transformer to take electronic functionality from the dense surface of the chip to the coarser pitched *outside world*,
- provide environmental isolation to the semiconductor, and
- provide an avenue for conduction of heat away from the device.

In the past, advances in silicon semiconductors had little or no affect on the electronic package design or process. New advances in silicon devices have driven increases in chip speed, device density (e.g., the number of transistors/mm²), and power dissipation. These changes, and those planned in the future, require larger numbers of interconnects to and from the devices, improved electrical performance, more effective thermal management, software systems to design and model the assembly, and different sets of materials to make these changes happen. Furthermore, the package is shrinking to a minimalist version of the products that traditionally have been called packaging. The chip and board

have become a conjugate packaging system that must provide the traditional levels of electrical, mechanical, thermal interfaces, and environmental protection, with the additional challenge of increased performance. Furthermore, the boundary between where the device ends and where the package begins has become blurred with passive devices (e.g., resistors, capacitors, etc.) in the package and redistribution layers that act as space transformers on the chip. Electronic packaging is becoming one of the greatest challenges in manufacturability, performance, and reliability in advanced electronics applications. With the tremendous growth of wireless telecommunication, radio-frequency (RF) applications are beginning to drive many areas of microelectronics traditionally led by the development of the microprocessor. An increasingly dominant factor in RF microelectronics is electronic packaging and the materials needed to create the package, because the package materials strongly affect the performance of the RF electronics. Many challenges also remain for packaging of microprocessors. These challenges include increased speed, numbers of input/output (I/O), decreased pitch, and decreased cost.

55.1 Package Applications

Electronic packages are divided into four levels, three of which exist beyond the integrated circuit.

Level 0: semiconductor chip level (integrated circuits). The materials are the semiconductor (Si, SiGe,

GaAs, etc.), metallization on the circuit (Al, Cu, Au) and interconnects on the die.

Level 1: chip in a carrier. The die is bonded (with a conductive, or nonconductive, adhesive or solder)

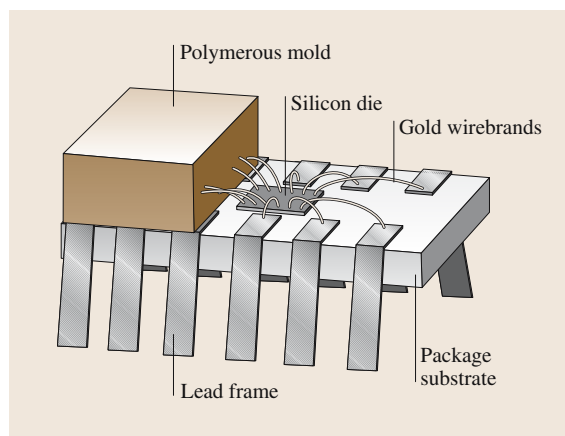


Fig. 55.1 Isometric section schematic that illustrates a peripheral package showing a dual-in-line package (DIP) that is wirebonded

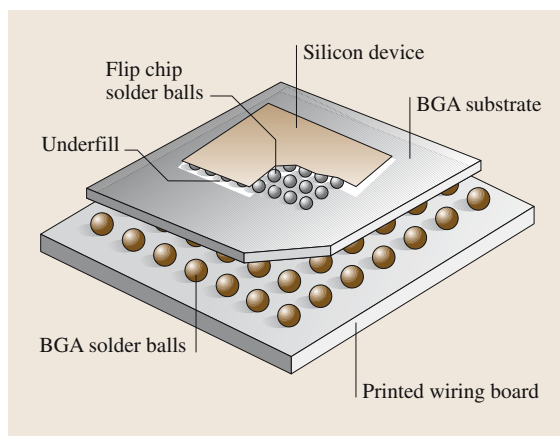


Fig. 55.2 Schematic illustration of an area-array package with a flip-chip device interconnect and a ball grid array package

into a package carrier (substrate or leadframe) and interconnected by either wirebonds or bulk conductive interconnects (flip chip) of solder or conductive adhesives. An example of a wirebonded package is shown in Fig. 55.1, while Fig. 55.2 shows a flip-chip package interconnect. The die is protected either by a lid or encapsulated with a polymer overmold.

Level 2: the chip carrier mounted to a board. The package is solder, or conductive adhesive, attached to

a circuit board. An example of a package mounted to a board using area array interconnects is shown in Fig. 55.2.

Level 3: board-to-board interconnects. The boards are interconnected to the final electronic system using friction interconnects, solder interconnects, or fiber-optic connectors.

55.2 The Materials Challenge of Electronic Packaging

Electronic packaging is arguably the most materials-intensive application today. The families of materials included in a package include: semiconductors, ceramics, glasses, composites, polymers, and metals. A list of the types of materials used in an electronic package are shown in Table 55.1. The processes required to assemble a package are equally varied: welding, soldering, curing, cold and hot working, sintering, adhesive bonding, laser drilling, and etching.

Each of these materials and how it is used in an electronic package could be the topic of an entire book. This

chapter is a high-level overview of the materials used in an electronic package. The focus is on the following classes of materials:

Metals are used as conductors in the package, primarily electrical but also thermal for power devices. This includes the thin metal interconnects on the integrated circuit made of Al, Cu or Au and the interconnects between the integrated circuit and the package that are either wirebond interconnects or solder joints. Metals are also used to act as heat sinks for power devices and as shields for RF applications.

Table 55.1 Examples of materials used in electronic packaging

Semiconductors	Si, SiGe, GaAs
Metals	Solders for interconnects (Sn–Pb, Sn–Ag, Sn–Ag–Cu, Sn–Au, Sn–Sb) Au wirebonds Cu leadframes (Kovar, CuBe, Alloy 42) Cu traces in substrates W, Mo traces in co-fired ceramics Ag, Au, Pd for thin/thick films on ceramics Ni diffusion barrier metallizations Al heat sinks
Ceramics	Al ₂ O ₃ substrates modified with BaO, SiO ₂ , CuO, etc. LTCC substrates Al ₂ O ₃ modified with low-temperature glass (e.g., PbO) SiN dielectrics Diamond heatsinks
Polymers	Epoxies (overmold) Filled epoxies (overmold) Silica-filled anhydride resin (underfills) Conductive adhesives (die bonding, interconnects) Laminated epoxy/glass substrates Polyimide dielectric Benzocyclobutene Silicones Photosensitive polymers for photomasks (acrylates, monomers, etc.)
Glasses	SiO ₂ fibers for optoelectronics Silicate glasses for sealing Borosilicate glass substrates Glass fibers for epoxy/glass substrates (FR-4)

Ceramics and glasses are used as dielectrics or insulators. Ceramics are used in devices as dielectrics to form capacitors and inductors. In the package itself, the ceramics are used as insulating materials for substrates that provide a structural base that electrically isolates lines and pads.

Polymers are used as insulators and, in a composite form, as conductors. As an insulator, polymers are encapsulants, underfills and substrates (note: in these applications, polymers are used as composites with silica or glass fillers). Polymers are also used as insulating adhesives to glue components to a substrate or board to provide mechanical strength. The addition of metal particles to the polymer can make it a conductive material as a conductive adhesive interconnect.

Composite materials are a mix of materials that can be tailored for either mechanical behavior improvements or thermal enhancement as an electrical conductor. Many of the composite materials used in packages are based on a polymer matrix, as noted in the polymer section above.

There is a wealth of general information on materials in electronic packaging. Some additional resources are found in [55.1–8].

55.2.1 Materials Issues in High-Speed Digital Packaging

The ability to decrease line width and feature size in semiconductor technology is decreasing below 90 nm. As the feature size shrinks, the function per unit area on the die increases. This increased functionality means either smaller die or more I/O per die. The reduced die size also reduces the available perimeter of the package for wirebond pads while increased functionality requires more I/O. The decreased perimeter area also drives wirebond capability down to 44- μm pitch with 10- μm gold wire. An attractive solution to this decreasing size is to use the entire surface of the chip (flip chip) rather than just the periphery (wirebond). Figure 55.3 shows a comparison of available I/O for a given chip size for standard, and state-of-the-art, wirebond and flip-chip pitch.

From an overall digital package perspective, the trend is to move from perimeter-bonded die and packages, to array packages, then to array die bonds in an array package that offers increased numbers of I/O and performance potential. This trend is shown in Fig. 55.4. Materials used in the package and substrate [e.g., ceramic (Al_2O_3), low-temperature co-fired ceramic (LTCC) and organic laminants] and the package style affect performance as well as area array ver-

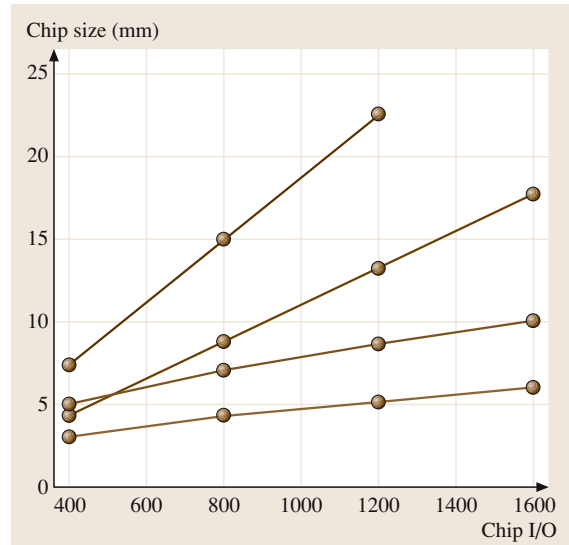


Fig. 55.3 Plot of bond pitch requirements as a function of I/O and die size for current wire bond (75 μm) and flip chip (250 μm) and state-of-the-art wirebond (44 μm) and flip chip (150 μm)

sus peripheral interconnects. Ceramic packages are the material of choice for hermetic high-performance applications. Ceramic is typically 96% Al_2O_3 with tungsten metallization. The ceramic is fired and sinters at high temperatures that precludes the use of Cu or Al metallization that would melt in the ceramic sintering process. Multilayer ceramic packages with thin-film metallizations also offer a thermal expansion coefficient similar

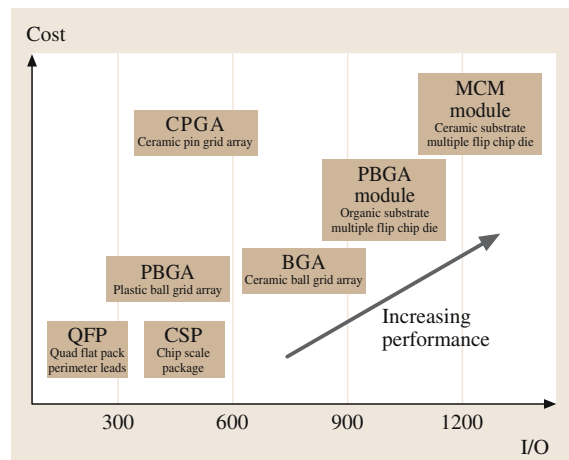


Fig. 55.4 Plot of package style as a function of cost and I/O

Table 55.2 Dielectric constants of packaging substrate materials

Material	Dielectric constant (ϵ_r)
Epoxy	5.0
FR-4	3.5–4.0
Other laminates	2.5–4.5
Kapton	3.1–3.5
Multilayer ceramic	5.0–9.5
Low-temperature co-fired ceramic (LTCC)	3.9–7.8

to the Si die ($6 \times 10^{-6} / ^\circ\text{C}$ for alumina ceramic versus $3 \times 10^{-6} / ^\circ\text{C}$ for Si) so strain that may arise during thermal cycling can be minimized. However, Al_2O_3 ceramic is expensive and has limited use in commercial applications. LTCC is growing in interest because it offers the hermeticity and thermal expansion advantageous of ceramic at a lower cost. LTCC is an alumina ceramic/glass composite that is fired at sufficiently low temperatures that Cu or Ag can be used as the metallization.

For the interconnects in the package, flip chip appears to provide substantial improvements in I/O and pitch but wirebonding will remain as a packaging solution because there is a great deal of capital invested in wirebond equipment that cannot be ignored.

In addition to smaller size driving finer pitch, increased signal speed will drive package requirements. Digital signal delays must be minimized. The total delay is a function of the total distance and the delay per unit length, which is a function of the transmitting medium's dielectric constant and is the square root of the dielectric constant (ϵ_r) times the free-space delay in vacuum (33 ps/cm). Therefore, a material with a high dielectric constant increases the delay. Table 55.2 shows a variety of dielectric constants used in packaging.

For minimal signal delays, an optimal dielectric substrate material is required. The signal length can also be shortened by changing from a wirebond solution to flip chip.

55.2.2 RF Packaging Materials Issues

High-speed and microwave circuits are defined for digital devices with clock speeds faster than 100 MHz and 0.1–100 GHz for analog circuits. For analog, these are also termed radio-frequency (RF) circuits. The microwave, or RF, circuit module is defined by microstrip elements composed of transmission lines and matching networks on a substrate with discrete components (resistors, inductors, capacitors and transistors)

Table 55.3 Coefficients of thermal expansion of a sampling of materials used in area-array electronic packages

Material	CTE ($10^{-6} / ^\circ\text{C}$)
Metals	
304 Stainless Steel	17.8
Ag	19.7
Al	23.5
Alloy 42	4.9
Au	14.2
Cu	16.8
Invar	1.6
Kovar	5.5
Mo	5.1
Ni	13–15
Solder: 63Sn–37Pb	25
Solder: 95Pb–5Sn	28
Ti	10
W	4.5
Ceramics	
AlN	4.3
Alumina (96%)	6.4
Alumina (99.5%)	6.5
BeO	7.8
BN	3.7
Fused silica glass	0.56
Quartz	13
SiC	3.8
SiN	3
Semiconductors	
GaAs	5.8
Si	2.7
Organic materials	
Epoxy resins	50–80
FR-4 (x–y plane)	15.8
FR-4 (z-axis)	80–90
Polycarbonates	50–70
Polyimide glass (x–y plane)	12–14
Polyimide glass (z-axis)	60
Polyimides	40–50
Polyurethanes	180–250
RTV (room temperature vulcanized) polymer	800
Sylgard	300

attached, or embedded, in the substrate. A monolithic microwave device has all the above elements integrated onto a semiconductor die. The semiconductor used for RF applications is typically GaAs because of its high resistivity and suitability for circuits that operate at high

frequencies. The following is a discussion of critical packaging materials issue related to RF devices and modules.

One of the most significant differences between RF and digital packaging is that the package is part of the RF circuit due to the interaction of the RF electric field with all adjacent conductors and insulators. This is one of the most significant challenges in the design of RF circuits. This will be further exacerbated as the trend moves from single-die RF packages to multiple die with passive components in RF modules. For example, the wireless-communication industry is striving toward a solution of a

phone in a package where the entire electronic functionality of a cellular phone can fit into a package with a size on the order of 1 cm^2 . Materials used in the package of these RF applications are significant for the performance and cost of the solution.

It is clear that the electronic package is a complex materials system that is driven by a variety of thermal, mechanical and electrical performance requirements. These requirements are also application dependent for both digital and RF packages. The remainder of this chapter is an overview of the key materials required for microelectronic packaging.

55.3 Materials Coefficient of Thermal Expansion

The coefficient of thermal expansion (CTE) is the length of increase of a macroscopic sample for a given temperature increase. The units of the CTE are length/length $^{\circ}\text{C}$. The CTE is a critical physical property of materials used in an area array package. A wide variety of materials with a wide variety of CTEs, such as metals, ceramics, and polymers are joined together

in an electronic package assembly. As processing (or in-use) temperatures change, the materials expand or contract to various degrees that could result in the formation of extensive and nonuniform strains in the package assembly. A set of materials typically used in electronic packages and their CTEs are shown in Table 55.3.

55.4 Wirebond Materials

Wirebonding is the process where a thin wire (that can be less than $25\text{ }\mu\text{m}$ in diameter) made of Au or Al is bonded to the surface of an integrated circuit, then to a pad or leadframe in the package (Fig. 55.1). Gold-ball thermosonic bonding is the typical method used to form these interconnects. The thermocompression bond is a weld between two metals where thermal energy (preheating of the capillary tool and wire), force and ultrasonic energy are imparted to the wire, causing it to melt on the surface and rapidly interdiffuse with the bond pad to form a joint. The integrated circuit pad's surface finish is typically Al or Au and the package bond pad finish is typically Au or Cu.

55.4.1 Wirebonds for Digital Applications

The challenges for decreased wirebond pitch are primarily process related to accurate machine control to move bonding heads to shorter distances ($44\text{-}\mu\text{m}$ pitch) with increased accuracy. However, there are also materials challenges. The metallurgical challenges of wirebonding Au to Al have been well addressed. The classic, early failures were dominated by *purple plague* and the definitive work in this area is by *Philofsky* [55.9]. Purple

plague resulted when the Au of the wirebond extensively reacting with Al to form brittle AuAl_2 that resulted in bond failures. (Purple plague derives its name from the purple tinge of the AuAl_2 intermetallic.) This was addressed by reducing the wirebond process temperature to below 300°C , which reduces the extent of AuAl_2 formation. Current wirebond failures are the result of surface impurities or corrosion. These are addressed by plasma or ultraviolet (UV)–ozone cleaning of the surface prior to wirebonding.

Wirebonding problems have also been caused by plating impurities in Au bond pads. Impurities in the Au diffuse to the molten front that occurs during wirebonding and concentrate in a plane in the bond. If there are a sufficient number of impurities, they precipitate and act as sinks for vacancies that can become voids and lower the bond strength. The impurities have been identified as Ni, Fe, Co and B [55.10]. The contaminants are present in the Au film as a result of the plating process. This can be addressed by optimizing the plating process to form pure metallic Au.

With finer-pitch wirebonding comes the requirement for thinner Au wire, approaching $10\text{ }\mu\text{m}$ in diameter. This thin wire poses a materials challenge. The wire

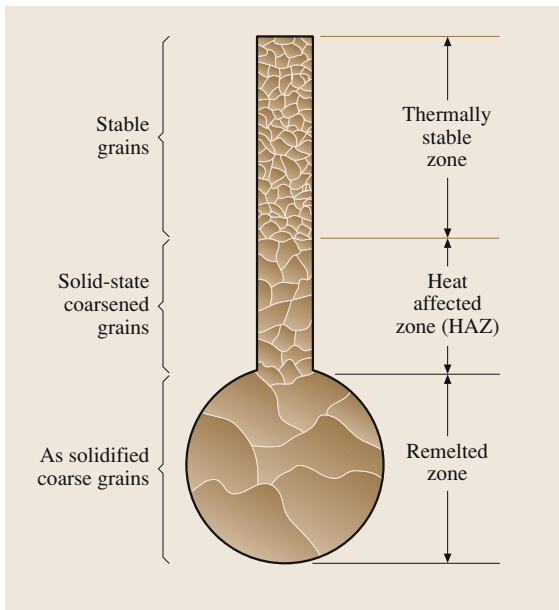


Fig. 55.5 Schematic illustration of the grain size in a Au wirebond. (After [55.11])

must have sufficient ductility to deform to the required wirebond shape but retain sufficient strength not to deform or move during the molding process (wire sweep). The Au alloy must be optimized for these opposing requirements. Au wire is typically stabilized with Cu and Ag dopants at levels below 100 ppm by weight. These dopants keep the grain size small to retain strength without deleteriously affecting ductility. As the Au wire decreases in width, the levels of dopants may need to be increased, reducing the Au to less than the current 99.9% level. Proposed dopants include Be and Cu. The critical region in the wirebond is above the ball and be-

low the wire, known as the heat-affected zone (HAZ). A schematic of the zone is shown in Fig. 55.5. This zone forms above the ball that was melted and cools to form a very large as-solidified grain size. The HAZ forms a large grain size while coarsening in the solid state as the molten ball is created. [The ball is formed under a flame or high current, solidifies, and is then placed on a Au or Al pad and with the addition of energy (pressure/heat/vibration) forms a solid-state ball bond.] The strength in the HAZ is reduced by the Hall-Petch relationship that states that strength is inversely proportional to grain size. The HAZ would be susceptible to potentially high levels of strain during molding and would be the most likely failure location. In thinner wires, an increased grain size is even more critical and alloying elements are needed to stabilize the structure and retain sufficient strength without affecting the ductility required to form the wirebond.

55.4.2 Wirebonds for RF

At RF frequencies, wirebonds affect electrical performance. At sub-RF frequencies, the wirebonds can be neglected as their resistance and inductance are in the noise with respect to performance. In RF applications, the wirebonds experience parasitic losses due to inductance with adjacent wirebonds, resulting in signal disturbance. RF-circuit designers typically add at least 1-nH inductance per wirebond but the effect can be variable and must be physically compensated by tuning the circuit. In effect, the wirebonds act as discrete inductors. The materials issues for RF wirebonds are the same as those described above for digital circuits but the effect of the wirebonds themselves is more significant in RF compared to digital and must be accounted for in the design of the device.

55.5 Solder Interconnects

For surface-mount and area array applications, the role of solder in the package is significant. In these advanced designs, the solder is an electrical interconnect, a mechanical bond and must often serve as a thermal conduit to remove heat from the joined device. The interconnects become more critical as chip size, chip carrier size, and the number of I/Os increase, while the solder joint size and cost decreases. Furthermore, some of the use environments are becoming increasingly severe. For example, some automotive electronics are being

placed in locations under the hood with temperature extremes that could range from -55 – 180°C . Portable devices, such as cellular telephones are expected to withstand severe shock environments caused by dropping a phone.

A solder interconnect consists of the solder alloy and the pads to which it is joined. The solder alloy is a low-temperature-melting metal that wets and reacts with metal pads or leads to form a metallurgical bond. During the wetting process the solder reacts with the

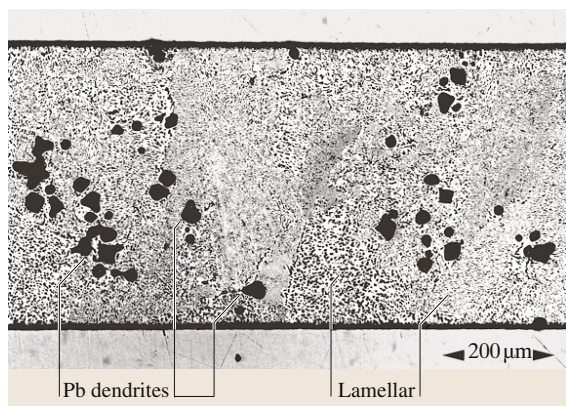


Fig. 55.6 Optical micrograph of the near-eutectic 63Sn–37Pb solder microstructure with the lamellar Sn- and Pb-rich phases and dendrites indicated

pad metallization, forming an intermetallic compound (IC) on the pad surface that continues to grow when the solder is in the solid state.

The most commonly used solder alloy for electronic interconnects is near-eutectic Sn–Pb with a composition of 63Sn–37Pb (by *wt%*). The microstructure of this alloy is shown in the optical micrograph in Fig. 55.6 and consists of Sn- and Pb-rich lamella. The lamella are oriented as cells after solidification with the slightly coarsened regions that separate the individual cells as the last part of the solder to solidify. This alloy melts at 183 °C. Pb-rich Sn–Pb alloys (such as 95Pb–5Sn or

97.5Pb–2.5Sn) are used for soldering temperature hierarchies as these alloys melt at temperatures in excess of 300 °C. The microstructure of Pb-rich alloys consists of a matrix of large Pb grains decorated with Sn precipitates. In flip-chip applications die are soldered to substrates with high-Pb solder. The substrates are soldered to a board with near-eutectic Sn–Pb. The Sn is the active components of these solder alloys and reacts with metallizations such as Cu or Ni to form intermetallic compounds.

Other solder alloys have found some use in electronic packaging, including Sn–Ag, Pb–In, Sn–Sb, etc. These alloys are used for a variety of applications including creating a soldering temperature hierarchy (higher or lower than Sn–Pb) or slowing the reaction rate with the metallization (e.g., In), or changing the mechanical properties of the joint. These alloys either form as a single-phase matrix with precipitates or as two-phase structures similar to eutectic Sn–Pb. The focus of this section, however, is on the most commonly used solder alloy, eutectic Sn–Pb.

The microstructure of the solder influences the mechanical behavior and reliability of the interconnects by governing the deformation and failure process. Typical deformation depends on the motion of dislocations and the growth and reconfiguration of grains and this is more sensitive to the structure of the solder than its chemistry. Therefore, depending upon processing conditions, the same solder alloy can respond differently to imposed stress and strain. Additionally, solders

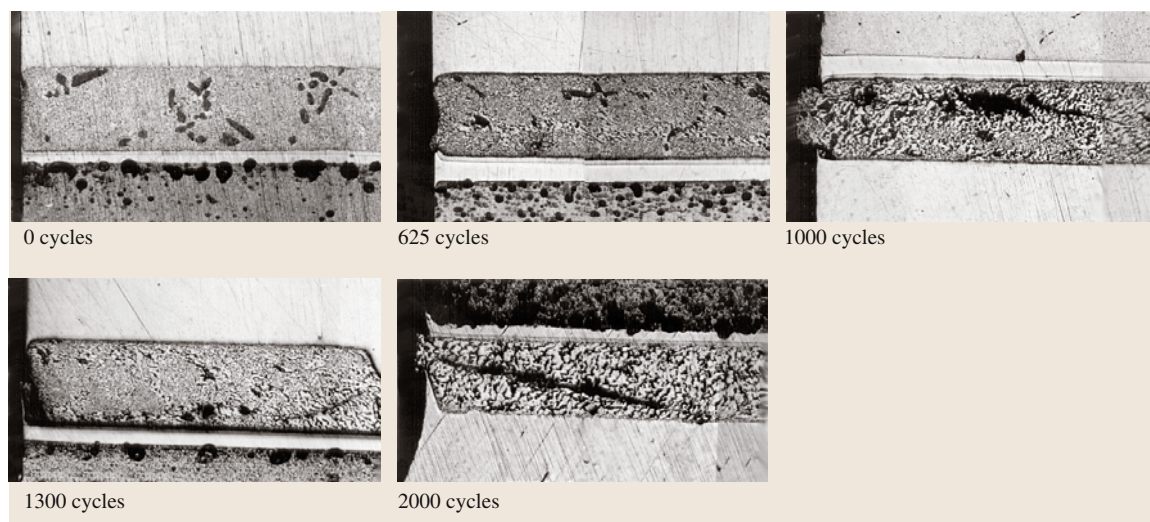


Fig. 55.7 Optical micrographs that show heterogeneous coarsening of the eutectic Sn–Pb microstructure at colony boundaries after thermomechanical fatigue

are low-melting-temperature materials that are used at a significant fraction of their melting point and are thermodynamically unstable. As the joint ages (at temperature and under strain) the microstructure changes and the mechanical properties and reliability change. The microstructure changes by grain (or phase) coarsening, which may be heterogeneous or homogeneous. Furthermore, the interfacial intermetallic grows in thickness and coarsens with time at temperature. All of this must be comprehended to determine the reliability of solder interconnects.

Thermomechanical fatigue occurs when materials with different CTEs are joined and used in an environment that experiences cyclic temperature fluctuations resulting in imposed cycling strain. Thermomechanical fatigue is a major deformation mechanism affecting solder interconnects in electronic packages. Even small temperature fluctuations can have a large effect, depending upon the joint thickness and CTE difference of the joined materials. The strain imposed on the solder joints follows the relation:

$$\Delta\gamma = \Delta\alpha\Delta Ta/h,$$

where $\Delta\gamma$ is the shear strain imposed, $\Delta\alpha$ is the difference in coefficient of expansion between the joined materials, ΔT is the temperature change, a is the distance from the neutral expansion point of the joined materials, and h is the thickness of the interconnect. After a critical number of thermal excursions, such as machine on/off cycles, solder joints experience fatigue failure. The type and magnitude of strains in solder joints under conditions of thermomechanical fatigue are often quite complex. For surface-mount applications, the strain is nominally shear. However, tensile and mixed-mode strains can occur due to bending of the chip carrier or board.

The combination of strain and temperature during thermomechanical fatigue has a large effect on the microstructure, and microstructural evolution of eutectic Sn–Pb solder joints. The microstructural evolution of 60Sn–40Pb solder as a function of the number of thermal cycles (-55 – 125°C) is shown in Fig. 55.7. The microstructure evolves through deformation that concentrates at the colony boundaries closely parallel to the direction of imposed shear strain, causing the cells to slide or rotate relative to one another. The structure within the cell boundaries becomes slightly coarsened relative to the remaining solder-joint microstructure and, thus are the weak links of the joint. Damage (in the form of defects or dislocations) is created at the cell boundaries at the low-temperature portion of a thermal

cycle. As the temperature rises, the deformation is annealed by recrystallization or stress-assisted diffusion, where material diffuses to regions of high stress. This results in coarsening of the Sn-rich and Pb-rich grains and phases in colony boundaries. The heterogeneously coarsened colony boundaries are weaker than the rest of the joint and any further deformation concentrates in the coarsened regions, resulting in further coarsening. Failure eventually occurs due to cracks that form in the coarsened regions of a joint. The first indications of impending failure are associated with cracking of coarsened Sn-rich grains in the heterogeneous region whose initial as-solidified grain size is in the submicron range. When cracks initiate during thermomechanical fatigue, the Sn-grains have grown to 5 – $10\text{ }\mu\text{m}$. Failure occurs when grains can no longer slide and rotate to accommodate the imposed strain, resulting in intergranular separation.

Lead-rich Pb–Sn alloys undergo thermomechanical fatigue behavior but show little evidence of microstructural evolution. These solders undergo intergranular failure caused by void coalescence and growth at Pb grain boundaries. The strain imposed during thermomechanical fatigue cannot be accommodated by large Pb grains, resulting in intergranular failure. However, Pb loses the work-hardening effect quickly so that damage in the form of cracks does not propagate easily.

Other solder alloys, such as the Pb-free Sn–3.5Ag eutectic-based solders, experience thermomechanical fatigue damage and failure at Sn grain boundaries. The microstructural evolution in these alloys tends to be phase coarsening with minimal grain-size coarsening. Sn–Ag–X alloys tend to have longer thermomechanical fatigue lifetimes than near-eutectic Sn–Pb solders.

Intermetallic compounds form between pad metallization and the active components of the molten solder (typically Sn). For Cu metallization, the Sn reacts to form Cu_3Sn and Cu_6Sn_5 intermetallics. For Ni, the Sn reacts to form Ni_3Sn_4 . After solidification, the intermetallic compounds continue to grow by solid-state diffusion. Over long periods of time, the intermetallic layers can grow to significant thicknesses ($> 20\text{ }\mu\text{m}$) and the solder–intermetallic interface may contain easy sites for crack initiation and propagation. Excessive growth also consumes the base metal, or finish, that can result in the loss of adhesion to the underlying metal that is not solder-wettable or create a plane of weakness owing to the stress generated from an intermetallic layer that is too thick. The metallized pad thickness must generally be greater than that consumed by the solder.

The transformation of solder-wettable coatings into intermetallics by solid-state reactions can also result in excessive intermetallic growth that degrades mechanical properties. The interfacial intermetallics are brittle and may fracture when strain is imposed, especially if the strain is tensile in nature. Solder-joint interfacial intermetallics are brittle because they typically have complex crystal structures with few crystallographic planes available to accommodate stress by strain relief, i. e., plastic deformation via a slip mechanism. The failures are characteristically brittle and occur through the intermetallic or at the intermetallic/solder interface under low-load conditions.

55.5.1 Flip-Chip Interconnects

Recent developments in under-bump metallurgy (UBM) and solder joints have resulted in lower cost and higher performance flip-chip interconnects. The UBM serves as an electrical, thermal, and mechanical interface between the silicon bond pads and the package substrate. The structure of the UBM is designed to adhere to the Al on the Si, act as a diffusion barrier between the solder and Si, and be a wettable surface to join to the solder. The most common UBM is the evaporated Cr/Cr-Cu/Cu/Au developed by IBM [55.12] for use with an evaporated high-Pb-content Pb-Sn alloy joined to a ceramic substrate. This UBM is expensive and new UBM systems with sputtered or plated metallizations have been developed. Plated metallizations are the lowest cost UBMs.

A significant change in flip-chip interconnects is the move toward a lower-melting-point solder alloy, such as eutectic 63Sn-37Pb (*wt%*) that can be processed below 220 °C as opposed to 350 °C for high-Pb solder. New materials on the die (low-*k* dielectric) and substrate (organic materials) require this lower processing temperature. The change in solder alloy also requires a change in the UBM. Traditional UBM systems are based on a thin Cu layer that dissolves into high-Sn solders, resulting in spalling and dewetting. Eutectic-solder UBM systems utilize a layer of Ni for solder wetting. The Ni is wet by Sn-Pb solder but reacts much more slowly than Cu. The addition of $\approx 9\%$ V to Ni reduces the ferromagnetic behavior of the Ni and enables the use of sputtering of the Ni. Electroless plating of Ni-9%P is the lowest-cost UBM and has the advantage that plating occurs only where desired on the metal pads on the silicon. To plate onto Al, the pads are zincated, then immediately plated with Ni. One issue with Ni-based UBMs is the intermetallic layer between the solder and

Ni (Ni_3Sn_4). Although very thin, this intermetallic layer can be brittle and the long-term reliability must be fully characterized and understood.

There are three solder deposition techniques available for flip chip. Traditionally, high-Pb solder is evaporated onto the UBM then reflowed to form metallurgical bonds. Evaporation is expensive because the masks must be regularly cleaned and there is excessive scrap solder. The composition of evaporated solder is difficult to control outside the high-Pb regime. Solder paste is a lower-cost alternative to evaporation and involves screening the paste onto the UBM then reflowing. Solder paste reduces waste, cleaning is inexpensive, and the paste comes in a variety of compositions, including eutectic Sn-Pb. The difficulty with paste arises in fine-pitch applications where the rheology of the paste makes it difficult to force into the required small holes in the screen. Solder may also be plated, which is inexpensive (no waste and easy cleaning) and has excellent fine-pitch coverage. Solder bumps at a pitch of 25 μm have been achieved by plating. The difficulty with plating is achieving ball and compositional uniformity.

The trend in flip-chip interconnect pitch is that the current 250- μm pitch will continue to shrink. The driving force for this pitch shrink is to satisfy the requirements for high-performance silicon devices. These requirements include a dramatic increase in the number of I/O due to increases in the number of signal lines and power requirements. Higher-power devices require more signal and ground lines and, to limit point sources of heat the power and ground interconnects should be spread evenly across the area array.

There are a number of materials and processing challenges associated with finer pitches. As the pitch shrinks, the methods to deposit the solder become more limited. Solder paste is very difficult to deposit using a silk-screen method at pitches below 150 μm due to rheological limitations of forcing a semi-solid (the paste) into small holes (the silk screen). Evaporation is difficult because developing a metal screen mask with the required tolerances is prohibitively expensive. Solder plating is still a good option but the solder must be very uniform across each die. Solder ball uniformity is critical because large variations between die could result in electrical opens, for small balls, and shorts, for large balls. At 250 μm , ball uniformity across a die is 10%, at 150 μm this decreases to 5% and the change in dimension tolerance is 12.5 μm down to 3.75 μm . An additional issue with a decrease in ball size is that the joint gap between the substrate and the die decreases to the point that it may

become very difficult to flow underfill completely under the die. At 100- μm pitch the gap between the die and substrate could be significantly less than 25 μm , below the limit of underfill flow. For these very-fine-pitch applications an alternative underfill technique will need to be developed because flow under the chip will not be possible. One alternative would be to deposit the underfill material on the wafer immediately after flip-chip solder bumping then underfill cure would simultaneously occur during solder reflow. New underfill materials and processes must be developed to implement this process.

55.5.2 Flip Chip for RF

Flip chip provides substantial improvements in off-circuit **RF** performance because the inductance is decreased as the ball height and shape is very consistent and predictable. Wirebond lengths can vary (typically $\pm 100 \mu\text{m}$), resulting in parasitic variations not observed in flip-chip interconnects. The number of I/O on an **RF** circuit is generally small so the pitch of the interconnects is generally quite large.

The die interconnects on GaAs are typically Au whereas on Si they are Al. A UBM is required for Si applications because solder does not wet Al. The solder would wet the Au of the GaAs but the Au layer is very thin and would dissolve the entire Au layer very quickly so a UBM is also required for GaAs. The UBMs available for Si can also be applied to GaAs but can be simpler because the adhesion of other metals to Au is easier to achieve than it is on Al.

The solder alloys for GaAs flip-chip bumps tend to be similar to that for Si but with a greater emphasis on Au-based alloys because of the desire to use a highly conductive material such as Au for very-high-speed applications. Eutectic Au–Sn, Ag and In alloys have been commonly used but the wetting of these solders is typically poor. For finer-pitch applications on GaAs (due to the small die size, not large numbers of I/O) other materials with good wetting are needed. Conductive adhesives have been explored for GaAs bumps. Work by Lin et al. [55.13] found no significant difference in electrical performance up to 2 GHz between a conductive adhesive and Au–Sn solder. The use of underfill for flip-chip packages can cause a major loss in **RF** energy because the high dielectric constant of the underfill polymer is greater than that of air. With a conductive adhesive localized to just the bond pads, not the entire surface of the die, this **RF** loss will not be observed.

55.5.3 Pb-Free

The electronics industry extensively uses Pb–Sn solder alloys in flip-chip applications as well as in many other interconnects in the electronic package. However, medical studies have shown that Pb is a heavy-metal toxin that can damage the kidneys, liver, blood, and central nervous system. Less than one percent per year of global Pb consumption is used in solder alloys for electronic products but electronics and electrical systems make up an increasingly larger fraction of landfills [55.14]. The issue of Pb leaching from landfills into the water table has raised alarm as a potential source of long-term contamination of soil and ground water. Concerns about the presence of Pb in the environment and potential exposure scenarios that could result in the ingestion of Pb by humans and wildlife have prompted a concerted effort to limit the use of Pb in manufactured products (notably gasoline, plumbing solders, and paint). International laws have recently been proposed to expand Pb control laws to limit or ban the use of Pb in manufactured electronics products. The most aggressive and well known effort is the European Union's Waste in Electrical and Electronic Equipment (WEEE) directive that proposes a ban on Pb in electronics by 2006. The Japanese Environmental Agency has proposed that Pb-containing scrap must be disposed of in sealed landfills to prevent Pb leaching. Electronics manufacturers have responded to these proposed bans in a variety of ways. Many companies have not taken a stance, hoping that legislation will not be enacted. Other companies have aggressively pursued solutions to the proposed bans and are using Pb-free products as a *green* marketing strategy. Extensive research on Pb-free solders has been published. A comprehensive review of the status of Pb-free solders, primarily focused on carrier-to-board (surface-mount and through-hole) interconnects, can be found in the literature [55.15–19]. A growing requirement is Pb-free solders for flip-chip interconnects.

One benefit of a Pb-free flip-chip interconnect is the reduction of Pb^{210} -created alpha-particle radiation. All mined Pb contains a small amount of radioactive Pb^{210} that decays and emits alpha particles. When an alpha particle enters an active element of the Si (such as a memory cell) it has sufficient energy to cause the stored charge to be released with the result of changing stored memory from a 1 to a 0 state. There is no permanent damage to the Si itself so this radiation-induced fault is termed a *soft error*. The alpha particles have a low energy that is dissipated over relatively short distances. However, due to their proximity to active

elements, the flip-chip solder interconnects have sufficient levels of alpha-particle radiation to induce soft errors in complementary metal–oxide–semiconductor (CMOS) technology, which become more critical as the cell size on the die is reduced [55.20, 21]. For Pb–Sn solders, one solution to alpha-particle radiation is to use elemental Pb that was mined many, many years ago where the majority of Pb^{210} has decayed. The source of this Pb is typically found as the ballast of shipping vessels that sank almost 2000 years ago and is relatively expensive. The elemental constituents of Pb-free solders (Sn, Cu, Ag, Bi, In, Sb), however, do not radioactively decompose so alpha-particle radiation is minimal.

Flip-chip interconnects are the electrical and mechanical connections between the semiconductor integrated circuit and the package [or board for direct chip attach (DCA)]. These interconnects are formed on the periphery or in an area array on the top surface of an active die. Flip-chip interconnects are formed by depositing solder onto a metallized Si wafer in the form of discrete balls, solder paste or by directly plating onto the pads on the wafer. The solder must wet and join to the pads on the Si devices so an under-bump metallurgy (UBM) is typically deposited on the Al or Cu pads on the Si. The UBM typically consists of a barrier metal (e.g., Ti or W) followed by a solder-wettable layer (e.g., Cu or Ni). The top layer of the Ni metallization is covered with a noble metal, such as Au, to prevent oxidation that

would inhibit solder wetting. The UBM also acts as a diffusion barrier between the Si and the solder and must be thick enough to withstand interactions (intermetallic formation) between the solder and UBM. Flip-chip interconnects are smaller (on the order of 100 μm in diameter) than surface-mount joints and are projected to have pitches that will shrink below 150 μm . Flip-chip interconnects have a unique set of requirements. These joints must be able to withstand a potentially high level of strain mismatch between Si and an organic substrate. Flip-chip technology has moved from ceramic packaging with high-Pb solder (97.5Pb–2.5Sn) to an organic package that requires lower-temperature reflow ($< 260^\circ\text{C}$). This can be accomplished by bumping the die with high-Pb solder then joining it to an organic board with eutectic Sn–Pb but this is a cost increase that is eliminated with a monolithic solder. The joints must withstand board-level reflow environments compatible with joining to organic substrates that, again, have a maximum reflow temperature of 260°C . The Pb-free solder must meet these requirements and perform at, or above, the level of performance of the Sn–Pb solder it is intended to replace. The flip-chip solder alloy is typically deposited on Si wafers either as solder paste stencil-printed on the defined UBM pads or by direct plating on the UBMs. The simplicity and low cost of plating the solder for flip-chip interconnects makes electrochemical deposition the most attractive choice for flip-chip bumping.

55.6 Substrates

The move to organic substrates is the focus of extensive development effort worldwide by both printed circuit board manufacturers and traditional suppliers that requires significant amounts of capital. Cost savings are expected by replacing ceramic substrates with organic substrates. One stated reason for moving to an organic substrate is based on the belief that, in volume production, organic materials are less expensive than ceramics due to the nature of printed circuit board processing. The organic substrate also offers an increase in electrical performance due to its lower dielectric constant (therefore, less capacitance and better speed) of 3.5 for organic versus 5.0 for ceramics.

Advanced substrates for flip chip must also provide increased wireability while delivering improved electrical performance with reliability levels at least equivalent to current surface-mount applications. The

board interconnect density for flip-chip substrates must accommodate the increasing density of off-chip interconnect, and cost-effective substrate capability that combines the necessary fine-line and micro-via features must be developed. Micro-vias are the metal-filled holes that provide a conduction path between copper lines in the multiple layers of an organic substrate. Micro-via capability must also scale with line width/spacing in order to provide the *via in line* structures which will be necessary to support bump pitch densification. At a pitch of 250 μm , micro-vias can be created using a photolithographic process. At finer pitch, laser drilling techniques are used to create the required small via size to support the smaller capture pads and facilitate the dense routing required for flip chip.

Materials with dielectric constants approaching 2.0 with coefficients of thermal expansion approach-

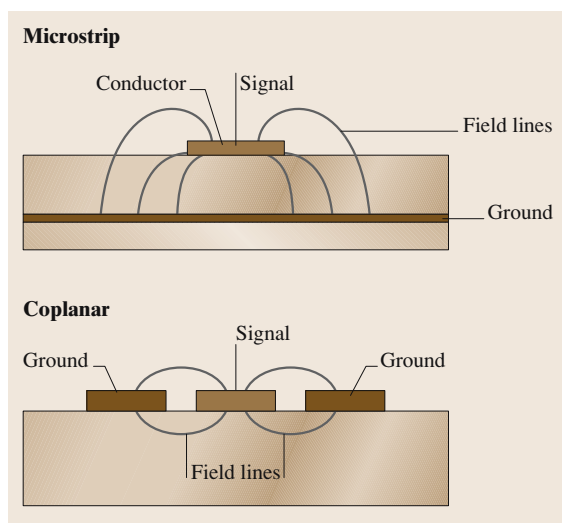


Fig. 55.8 Schematic illustration of RF electric-field lines in the microstrip and coplanar designs

ing 6.0 ppm/°C are necessary to meet fine pitch requirements. Furthermore, uniformity and flatness requirements must be improved on organic substrates to ensure uniform joint size and bonding across each die as the interconnect pitch decreases.

The change in substrate materials drives the need for a lower-melting-temperature solder to replace the high-Pb-content Sn–Pb alloys that are processed in excess of 300°C for ceramic substrates. The organic materials in a substrate are typically epoxy-based and char or burn at temperatures in excess of 250°C. The most common solder alloy candidate is near-eutectic Sn–Pb, which can be processed well below 250°C ($T_{mp} = 183^\circ\text{C}$). However, this could result in a situation where the flip-chip interconnect melts every time a subsequent solder reflow occurs. Multiple reflows of the solder interconnects accelerates intermetallic growth between the tin of the solder and the metallized pads for the UBM and substrate. One way to circumvent this issue is to deposit low-melting-temperature eutectic Sn–Pb solder on the substrate lands, or on the high-melting-temperature solder bump. Reflow to the substrate can be performed at eutectic Sn–Pb soldering temperatures but the mixing between the high-Pb-content and eutectic solder results in a higher-melting-temperature composition. Subsequent reflows, at eutectic Sn–Pb processing temperatures, will not remelt the *composite-alloy* flip-chip joint. This addresses the low-temperature processing requirements for organic substrates but does not address the low- κ -dielectric

low-temperature processing requirements because the high-Pb-content solder must still be reflowed on the silicon.

55.6.1 RF Substrate Materials

Optimum electrical performance is achieved in an RF module when the signal-carrying conductor is adjacent to the ground plane to contain the electric field. In microstrip designs the conductor is above the ground plane, in coplanar designs the conductor is sandwiched between ground planes, as shown schematically in Fig. 55.8. These designs affect the materials in the package. The impedance of the lines is desired to be 50 Ω (which matches the radiation impedance of the antenna to air). The impedance is affected by the conductor material, typically Au, and the dielectric material [Si, SiGe, GaAs, for the die and LTCC, Al_2O_3 or epoxy (organic) for the substrate].

Like its digital counterpart, RF packaging is moving away from the Al_2O_3 ceramic substrates used in RF applications to the lower-cost solution of LTCC and organic materials. The metallizations used are Cu on the interior of the substrates and Au as the top-layer metallization to improve RF performance.

Organic substrates offer a lower-cost solution. This substrate is built in a laminate structure with punched, or laser-drilled, holes that are subsequently plated for interlayer via interconnects. The laminant layers are variations of epoxy and glass optimized for low dielectric constants. LTCC substrates are also laminant structures of sheets of LTCC with laser-drilled and plated vias. Recently, LTCC materials have been optimized to not shrink in the x – y plane during the sintering process (z -direction shrinkage, however, is significant). The LTCC laminants are extremely thin, compared to organic, and offer the opportunity to build complex RF substrates.

An additional trend in RF substrate technology is to move the discrete functions of RF module circuits from surface-mounted components to embedding them in the substrate. These embedded functions are typically capacitors and inductors and occasionally resistors. This can be done because the substrate material is dielectric and metal, which are the basic components of passive devices. The advantages of embedding components are reduced size, reduced cost and improved performance. Size can be reduced because surface-mount components take up a great deal of the surface of the substrate. By embedding these components, they are thinner and the area is reduced. Cost is reduced because the embedded

passives are formed as part of the laminant layers by patterning the metal/dielectric and is essentially cost-free because the patterning would be done regardless. Performance can be increased through the smaller size of the module. The embedding of components in substrates does offer the opportunity for improved materials. Per-

formance could be improved by finding means to alter the materials properties of regions of the laminant layers to optimize the electrical performance of discrete components that are to be embedded (i. e., *seed* regions of the substrate for optimal capacitance, inductance and resistance).

55.7 Underfill and Encapsulants

55.7.1 Underfill

Flip-chip technology demands that the space between the chip and the substrate be filled with a dielectric organic material to help mechanically interlock the chip to the substrate. Most underfill materials available are silica-filled anhydride resin polymers. The filler provides the strength in the composite polymer and the resin bonds with the die and substrate. Without the underfill the difference in coefficient of thermal expansion between the chip ($3 \text{ ppm}/^\circ\text{C}$) and the substrate ($\approx 17 \text{ ppm}/^\circ\text{C}$) would quickly result in fracture of the solder interconnects when variations in temperature occurred. New underfills must be developed to accommodate increased bump density and shrinking of bump height. As bump density increases and pitch decreases, the underfill must flow into smaller spaces. A $250\text{-}\mu\text{m}$ pitch requires an underfill with a thickness of $110\text{-}\mu\text{m}$. At a $150\text{-}\mu\text{m}$ pitch, the dimension between the chip and substrate will be less than $50\text{-}\mu\text{m}$ but 100% fill of underfill is still required. Traditional flow dispensing and curing may not be suitable in these thin spaces due to a lack of capillary driving force (current underfill materials flow through a minimum thickness of $75\text{-}\mu\text{m}$). For very thin fill spaces, wafer-level deposition of the

underfill needs to be developed. This wafer-level deposition involves spinning on of an uncured polymer to a uniform thickness slightly less than the bump height. The wafer would then be diced and die-mounted on substrates. Reflow would form the solder joints and cure the underfill, simultaneously. The filled polymer must be spun to a uniform thickness of $25\text{-}\mu\text{m}$ or less, and bond well with silicon and substrate materials, and also be compliant.

55.7.2 Encapsulation

Encapsulation of electronic packages is required to protect the electronic components from environmental degradation such as moisture contamination, mechanical damage and must also provide electrical isolation. The encapsulation material is a polymer that flows over and covers the electronic components, and then cures into a solid. For electronic packaging, the most common method for applying the encapsulation is transfer molding. Figure 55.9 is a schematic illustration of the transfer mold process. The transfer mold compound material is typically a thermoset cross-linking polymer. The polymer resin portion forms 25–35% of the total mold compound (by weight) and consists of epoxy resin, a hardener and a curing agent. The mold compound material is filled with powders of silica (SiO_2), alumina (Al_2O_3) or glass (65–75% of the total weight) to reduce the thermal expansion coefficient of the polymer and add strength. Amorphous silica is used for minimum thermal expansion coefficient but this results in a decrease in thermal conductivity. Crystalline silica improves thermal conductivity but is abrasive to the molds themselves. Small amounts of pigments (to color the mold compound), accelerants, flame retardants, mold-release agents, getterers, and antioxidants are also added to the final mold compound. Pellets of the mold compound are placed into a heated cavity at the top of the mold and pressure is applied by a ram that pushes the liquefying materials into the mold. The material

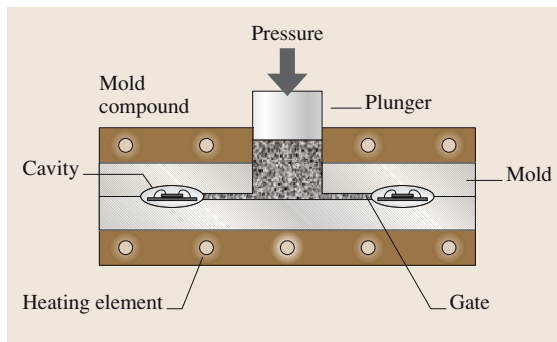


Fig. 55.9 Schematic illustration of the transfer mold process

is forced through channels and through a constriction (called the gate) that provides back-pressure and regulates the flow rate of the molten mold compound. After passing through the gate, the mold material flows across the component being encapsulated. The heat is turned off and upon cooling the polymer cross-links, with the use of a curing agent (typically an amine) that is included in the mold compound. The mold compound material has good adhesion, good stability with regards to temperature, chemicals, and the environment. It also cures rapidly, has low permeability to moisture and high dielectric strength.

The transfer mold process has the advantage that part dimensions are very well controlled. The process and materials are low cost and high-volume manufacturing is possible.

An issue with the mold compound encapsulation processing is wire sweep, where the force of the viscous encapsulant can push bond wires to short with one another. This has been addressed with optimized filler size and shape and mold temperature and pressure. After molding, the difference in thermal expansion between the cured mold compound and the rest of the package (the substrate, leadframe, semiconductor device, etc.) can result in considerable stresses at the interfaces. Furthermore, the CTE of mold compounds changes when the material goes above the glass-transition temperature, T_g . T_g is the temperature at which the polymer changes from a rigid structure to a softer, more glass-

like structure. Below T_g , the CTE of typical mold compounds is $20 \times 10^{-6} / ^\circ\text{C}$ but above T_g this can increase to $60\text{--}70 \times 10^{-6} / ^\circ\text{C}$. The T_g of mold compounds varies but is usually $175\text{--}200^\circ\text{C}$ below the solder processing temperatures, which are typically in excess of 220°C . After cooling from mold processing, the mold compound forms a compression seal with the materials that expand less (leadframe, substrate, semiconductor). However, upon heating to solder process temperatures, the stress is reversed to tensile and can result in interfacial cracking, or delamination. The interfacial cracks provide a conduit for the ingress of water or ionic contaminants that could result in corrosion of metals in the package (particularly Al bond pads). This issue must be addressed by optimizing the mold compound to the package application (materials and size) to minimize strain and to maximize interlocking mechanisms between the mold compound and the substrate/leadframe.

Mold compound encapsulation is generally detrimental to RF device performance. The encapsulant materials are nonuniform and degrade performance because the electric field generated by the RF circuit in the encapsulant is nonuniform, resulting in lossy behavior. Many high-performance RF packages are designed with a cavity and are made out of ceramic with a metal lid to form the cavity. However, plastic encapsulation is required to achieve the required low-cost reliability of the package, so a trade-off must be made for performance, reliability and cost.

55.8 Electrically Conductive Adhesives (ECAs)

Electrically conductive adhesives (ECAs) are composite materials consisting of a dielectric curable polymer and metallic conductive particles. ECAs are low-temperature alternatives to solder alloys. The polymer is an adhesive material that reacts chemically with metals to form a bond. The metallic particles in the adhesive form a network in the cured joint that forms a conduction path in area-array applications.

55.8.1 Adhesive Polymers

The polymer portion of the adhesive is available in two forms: thermoplastic and thermosetting. A schematic illustration of these two types of polymers is shown in Fig. 55.10. A thermoplastic polymer is a linear chain of linked monomers (nominally, hydrogen-carbon bonds). After curing, thermoplastic polymers becomes

more rigid, but remain flexible. As the temperature is increased, these polymers essentially become molten. This provides a method to repair or replace components with thermoplastic joint area-array joints. Thermosetting materials are similar to thermoplastics, except that monomer chains become cross-linked in three dimensions during the curing stage. The number of cross-links determines the rigidity and the glass-transition temperature (T_g) of a thermoset polymer. In general, the more cross-links, the higher the T_g . A curing agent (e.g., a hardener) can be added to thermoplastic polymers to cause the cross-linking reaction in these materials as well. Curing can be performed using UV light, heat, or catalysts. Thermosetting materials are harder than thermoplastics, but can be processed and cured below their T_g and do not become molten above the T_g . However, the curing operation is not reversible, mak-

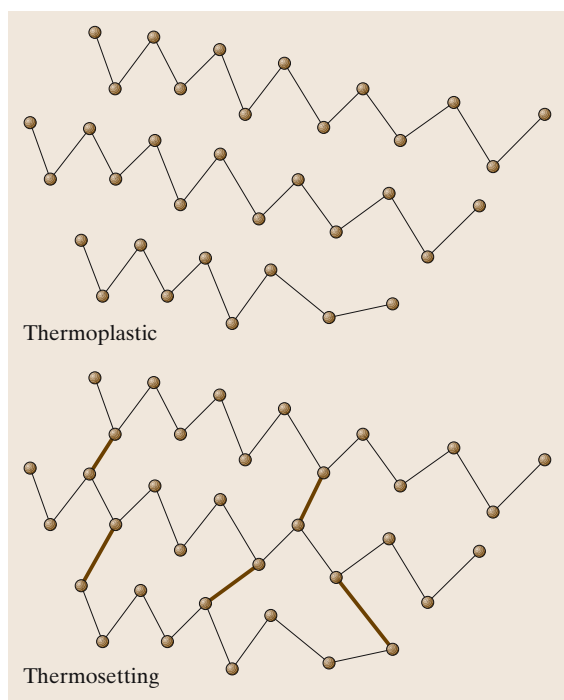


Fig. 55.10 Schematic illustration of polymer bonds for thermoplastic (linear bonds) and thermosetting (cross-linked bonds) electrically conductive adhesives

ing repair difficult. Examples of thermosetting materials include epoxies and acrylics. ECAs are typically made using either thermosetting or thermoplastic polymers, or a combination of the two.

55.8.2 Metal Fillers

A number of metals have been used to provide the electrical conductivity of ECAs. The metals are typically in the form of flakes, plates, rods, fibers, or spheres. The size of the metal particles is 5–20 μm . For the ECA to have suitable strength and electrical conduction requires a trade-off in metal filler content. The higher the metal content, the better the electrical conduction, but the poorer the strength of the adhesive. The metal content depends upon whether the adhesive is an isotropic or anisotropic conductor (described below), but does not usually exceed 40% by volume.

The metals most often used in ECAs are silver, gold, nickel, copper or variations of these (such as gold on nickel, silver on glass, or nickel on polymers). Silver-filled ECAs are the most common, because the cost of the metal is moderate and it has good electrical conductivity

and low reactivity with oxygen. Gold has better physical properties than silver but the cost is prohibitive for most applications. Nickel is less expensive, but also has lower conductivity than either Au or Ag and has been found to corrode when aged in a humid environment, resulting in poor adhesion between the nickel and the epoxy matrix. Copper fillers oxidize rapidly and also delaminate from the polymer matrix.

55.8.3 Conduction Mechanisms

The mechanism of conduction in ECAs is hypothesized to occur via percolation, where current travels through a three-dimensional matrix of contacting particles. There must be a sufficient quantity of particles present in the adhesive to provide a suitable conduction path. The bulk resistivity of 60Sn–40Pb solder is $15 \times 10^{-6} \Omega \text{ cm}$, while that of ECAs is 7×10^{-5} to $5 \times 10^{-4} \Omega \text{ cm}$ [55.22]. Therefore, the conduction of solder is 5–33 times better than ECAs. This difference in conductivity is due to the fact that only a portion of the joint cross section is available to provide the percolation path in ECAs whereas the entire cross section of solder joints conducts electricity. The loss of electrical conductivity does not severely affect second-level interconnects but could be an issue for first-level interconnects where good conductivity and switching speed may be required. For the current to pass from one metal particle to another, it must pass through, at a minimum, the oxide that forms on the metals or, at worst, some distance through the polymer matrix. It has been proposed that contact resistance is overcome by tunneling or bulk conduction through the semiconducting metal-oxide layer. However, measured electrical conduction has been determined to be better than any of the mechanisms described above [55.23]. It is possible that, in silver-loaded ECAs, silver diffuses out of the particles into microcracks in the polymer matrix, thereby creating additional conduction paths. To date, the details of the conduction mechanisms remain unresolved.

55.8.4 Isotropic Versus Anisotropic Conduction

ECAs are available in two types of conductors, isotropic and anisotropic. In isotropic adhesives, the conduction path is uniform in all directions with a filler-metal content sufficiently high to create a percolation path for the conductor. Anisotropic ECAs (or AECA, also called z-axis conductors) only have one direction of conduction, useful for fine-pitched applications. In general,

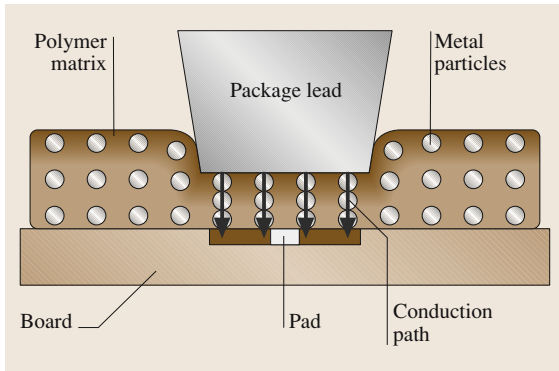


Fig. 55.11 Schematic illustration of an anisotropic electrically conductive adhesive. The adhesive is compressed, causing the conductive particles to contact one another in the z -axis and form an electrical path between the terminal pads. Outside the pad region, the metal particles are separated by the polymer and the AECA acts as a dielectric

anisotropic ECAs are prepared by dispensing metal particles into the polymer matrix at a level far below the percolation threshold. Pressure is imposed on the contact so that the particles come into contact with one another in one direction (parallel to the direction of imposed pressure). In the areas where no pressure is imposed, there is no conduction path and the adhesive is electrically insulated. This is illustrated in Fig. 55.11. In area-array applications anisotropic ECAs can be applied to the board but conduction will only take place between package and mating board pads due to the compression resulting from the cured adhesive, causing particle-to-particle contact between mating pads. No metal particle contact is made beyond the terminal pads, thus the adhesive is an insulator in those areas.

The terminal pad surface to which ECA or AECAs are to be joined must be oxide-free for the adhesive to form a cohesive joint. Bare copper typically provides very poor adhesion [55.24] because adhesives do not wet or react well with metal oxides, and in this

way adhesives are similar to solder alloys. To overcome this problem, copper can be plated with nickel ($\approx 0.5\ \mu\text{m}$) and a thin layer of gold ($\approx 500\ \text{\AA}$). Gold is an excellent bonding surface because it is virtually free of surface oxide and provides a diffusion barrier for nickel. Unlike solders, ECAs do not allow the luxury of using flux to remove oxides during processing so oxide-removal surface treatments must precede dispensing the adhesive. The stability of ECA joints is directly related to the quality of the bond interface with terminal-pad metallizations. In general, the less surface oxide present, the better the electrical stability. If surface oxide is present, there can be de-adhesion at the interface, which results in a reduced ability to carry electrical signals. The bonding of ECAs to other polymers, such as printed wiring board surfaces (FR-4 or polyimide) is excellent. Area-array component requirements are more stringent in the case of ECA joints compared to solder bumps. Solders only wet metal surfaces, so special processing steps are not needed to define solder-joint locations. For ECAs, the method of depositing the ECA must be precise (or an anisotropic ECA used) so that electrical shorts do not occur and this is exacerbated in area-array fine-pitch applications. Also the interconnects cannot easily be inspected because they are hidden beneath the joined component.

55.8.5 Rework

The ability to perform rework on conductive adhesive interconnects may limit the use of these materials for many applications. Thermoplastic adhesives can be reworked by heating and melting the joint, allowing components to be separated from the assembly. Thermosetting bonds are not reversible but can be softened by heating well above T_g and then removed. Some acrylic adhesives can be cleaved at room temperature and physically removed, although the brittle nature of the acrylics that allows for removal may also represent too great a joint reliability risk.

55.9 Thermal Issues

55.9.1 Thermal Issues in Digital Packaging

Increasing circuit and component density creates thermal complications that must be resolved by the packaging solution. Excessive temperatures in the de-

vices result in degraded circuit performance or reliability issues. The thermal performance can be improved by using packaging materials with high thermal conductivity.

The operating temperature of Si directly impacts on switching speed and timing. In the operating temperature

range of 50–100 °C, a 5 °C drop in operating temperature results in a 1% increase in CMOS switching speed. From a reliability perspective, each 10 °C increase in operating temperature results in a twofold increase in the die failure rate. Therefore, packaging materials and processes that improve thermal performance are very important.

New materials is the area where the most significant improvements in thermal performance is possible. Improved thermal interface materials (greases, adhesives, tapes, phase-change materials) will improve heat flow out of the package. Improved adhesives and compliant sheets inside the package will improve heat dissipation from the die and can better accommodate CTE mismatch strains within the package.

Composite materials for package lids, heat spreaders and heat sinks are needed. Currently, these components face a dilemma of conflicting materials properties as good materials with good thermal conduction (e.g., Al) have very high CTE (e.g., $\text{Al} = 25 \times 10^{-6} / ^\circ\text{C}$), which can result in poor reliability. Furthermore, most low-cost thermal conductors are also electrical conductors, which makes their use in dielectric packages problematic. The development of composite materials with anisotropic conductivities can provide excellent thermal conduction with matched CTE and the potential for good electrical isolation. At the die level, the use of low-cost thin-film diamond must increase. Diamond has excellent thermal conductivity and isolation but is currently too expensive for extensive use in microelectronic packaging.

55.10 Summary

Materials are one of the most important aspects of microelectronic packaging. There is a need for new materials to be developed to meet the ever-changing requirements of high-speed digital and RF applications. Considerable resources are required to develop and characterize these materials. These materials must also be tested to determine their compatibility with assembly processes and

55.9.2 Thermal Issues in RF Packaging

RF die can generate a great deal of heat when operating at high frequencies. This is particularly an issue with GaAs die because it has poor thermal conductivity (30 W/mK). The generated heat must be dissipated through the package. The preferred method to remove heat from wirebonded GaAs die is to thin the GaAs wafer by back-grinding to 100–125 μm . The thinner the die, the easier it is to transmit heat through the GaAs. Through-die vias are also used to help dissipate heat. The vias are formed by laser-drilling holes through the GaAs then plating the side walls with Au then thinning the wafer. Thinning the wafer not only benefits heat dissipation but also reduces the distance from the conducting layer on the surface of the die to the ground plane on the die backside, which improves impedance match in microstrip designs.

The problems with through-die vias and wafer thinning are cost and die breakage. Laser-drilling and thinning process increase the cost of the die. Furthermore, GaAs is a very brittle material and the presence of vias and thinning exacerbates the potential for cracking.

One method to reduce the thermal and mechanical problems associated with thermal issues on GaAs is to use flip-chip interconnects. For flip chip, the bumps act as electrical, mechanical and thermal interconnects. The bumps are on the front side of the die so full-thickness die can be used. The challenge with flip chip as a thermal conductor is to put sufficient numbers of bumps on the die to dissipate the heat.

device performance and to meet industry standards. The fundamental materials characteristics that must be understood, depending upon the material, include: the dielectric constant, dielectric loss, conductivity, resistivity, moisture absorption, glass-transition temperature, strength, time-dependent deformation (creep), and fracture toughness.

References

- 55.1 D.R. Frear, W.B. Jones, K.R. Kinsman: *Solder Mechanics: A State of the Art Assessment* (TMS, Warrendale 1991)
- 55.2 *Area Array Packaging*, ed. by K. Puttlitz, P. Totta (Kluwer, Dordrecht 2001)
- 55.3 *High Temperature Electronics*, ed. by P.A. McCluskey (CRC, Boca Raton 1996)
- 55.4 J.H. Lau: *Ball Grid Array Technology* (McGraw-Hill, New York 1995)
- 55.5 G. Harmann: *Wire Bonding in Microelectronics Materials, Processes, Reliability, and Yield* (McGraw-Hill, New York 1997)
- 55.6 R.J. Klein Wassink: *Soldering in Electronics* (Electrochem. Publ., Ayr, Scotland 1989)
- 55.7 C.A. Harper: *Electronic Packaging and Interconnection Handbook* (McGraw-Hill, New York 1991)
- 55.8 *Electronic Materials Handbook, Vol. 1: Packaging*, ed. by M.L. Mingos (ASM-Int., Materials Park, OH 1989)
- 55.9 E. Philofsky: Intermetallic formation in Au-Al systems, *Solid State Electron.* **13**, 1391–99 (1970)
- 55.10 C. Horsting: *Purple Plague and Au Purity, Proceedings 10th Annual IRPS* (IEEE, Westmoreland, NY 1972) pp. 155–8
- 55.11 L. Levine, M. Sheaffer: Wirebonding strategies to meet thin film packaging requirements – Part 1, *Solid State Technol.* **36**, 63–70 (1993)
- 55.12 L.S. Goldman: Geometric optimization of controlled collapse interconnects, *IBM J. Res. Dev.* **13**, 251 (1969)
- 55.13 J.-K. Lin, J. Drye, W. Lytle, T. Scharr, R. Sharma: *Conductive Polymer Bump Interconnects, Proceedings of the 46th Electronic Components Technology Conference* (IEEE, Piscataway, NJ 1996) pp. 1059–68
- 55.14 N.C. Lee: *Pb-free soldering – Where the world is going*, *Adv. Microelectron* (IEEE, Piscataway, NJ 1999) p. 29
- 55.15 J. Glazer: Microstructure and mechanical properties of Pb-free solder alloys for low-cost electronic assembly: A review, *J. Electron. Mater.* **23**, 693–700 (1994)
- 55.16 J. Glazer: Metallurgy of low temperature Pb-free solder for electronic assembly, *Int. Mater. Rev.* **40**, 65–93 (1995)
- 55.17 M. Abtew, G. Selvardery: Pb-free solder in microelectronics, *Mater. Sci. Eng.* **27**, 95–141 (2000)
- 55.18 H.K. Seelig, D. Suraski: *The Status of Pb-free Solder Alloys, Proc. 50th Electron. Comp. Tech. Conf.* (IEEE, Piscataway, NJ 2000) pp. 1405–9
- 55.19 K.G. Snowden, C.G. Tanner, J.R. Thompson: *Pb-free Soldering Interconnects: Current Status and Future Developments, Proc. 50th Electron. Comp. Tech. Conf.* (IEEE, Piscataway, NJ 2000) pp. 1416–19
- 55.20 Z. Hasnain, A. Ditali: *Building-in reliability: Soft errors – a case study, Ann. Proc. Reliab. Phys.* (IEEE, Westmoreland, NY 1992) pp. 276–80
- 55.21 M.W. Roberson, P.A. Deane, S. Bonafede, A. Huffman, S. Nangalia: Conversion between standard and low-alpha Pb in solder bumping production lines, *J. Electron. Mater.* **29**, 1274–7 (2000)
- 55.22 H.L. Hvims: Conductive adhesives for SMT and potential applications, *IEEE Trans. Components Hybrids Manuf. Technol.* **18**, 284–91 (1995)
- 55.23 L.C. Li, H. Lizzul, I. Kim, J.E. Sacolick, J.E. Morris: Electrical, structural and processing properties of electronic conductive adhesives, *IEEE Trans. Components Hybrids Manuf. Technol.* **16**, 843–51 (1993)
- 55.24 D.D.L. Chang, J.A. Fulton, H.C. Ling, M.B. Schmidt, R.E. Sinitiski, C.P. Wong: Accelerated life test of z-axis conductive adhesives, *IEEE Trans. Component Hybrids Manuf. Technol.* **16**, 836–42 (1993)

Acknowledgements

1 Perspectives on Electronic and Optoelectronic Materials by Tim Smeeton, Colin Humphreys

Many thanks to Dr Ron Broom for his helpful comments on this manuscript.

A.5 Defects in Monocrystalline Silicon by Wilfried von Ammon

The author is greatly indebted to W. Haeckl, E. Dornberger, D. Gräf and R. Schmolke for many helpful discussions and for providing some of the figures.

A.7 Photoconductivity in Materials Research by Monica Brinza, Jan Willekens, Mohammed L. Benkheldir, Guy J. Adriaenssens

The authors are grateful to the *Fonds voor Wetenschappelijk Onderzoek – Vlaanderen* for its financial support of their research.

A.9 Charge Transport in Disordered Materials by S. D. Baranovskii, O. Rubel

The authors are indebted to numerous colleagues for stimulating and enlightening discussions. Among those are Boris Shklovski (University of Minnesota), Alexei Efros and Michael Raikh (Utah University), Hellmut Fritzsche (Chicago University), Peter Thomas, Walther Fuhs and Heinz Bässler (Philipps-University Marburg), Igor Zvyagin (Moscow State University) and many other colleagues. Financial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

A.11 Ionic Conduction and Applications by Harry L. Tuller

Support from the National Science Foundation (Grant Nos. DMR-0243993 and ECS-0428696) and ARO-MURI under grant DAAD-0101-0566 for topics related to this work are highly appreciated. In assembling this work, I drew on earlier journal and proceedings articles published by myself or in conjunction with colleagues. In particular, I wish to acknowledge my collaborator in Ref. 2, Prof. P. Knauth of the Université de Provence, Marseille, France.

B.17 Structural Characterization by Paul D. Brown

As ever, there are many people one wishes to acknowledge for their involvement in the growth, processing

and underpinning characterisation research programmes drawn from to illustrate this chapter. University of Nottingham: with thanks to Tom Foxon, T.S. Cheng, Sergei Novikov and Chris Statton for the provision of MBE GaN samples and supporting XRD analysis; and to Mike Fay for GaAs CBED patterns. University of Cambridge: with thanks to Colin Humphreys for provision of instrumentation; Chris Boothroyd for EDX and HAADF data on the SiGe samples; Michael Natusch for GaN EELS data; Robin Taylor for RHEED stage development; David Tricker for the Si-doped GaN micrograph; and Yan Xin for the GaN images used for dislocation analysis. University of Warwick: with thanks to Richard Kubiak and E.H.C. Parker for supplying SiGe/Si samples. Polish academy of Sciences, Warsaw: with thanks to Jan Weyher for homoepitaxial GaN samples. With thanks also to the EPSRC for funding support.

B.19 Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications by S.O. Kasap

The authors thank NSERC for financial support.

B.20 Electrical Characterization of Semiconductor Materials and Devices by M. Jamal Deen, Fabien Pascal

The authors are very grateful to Drs. O. Marinov and D. Landheer for their careful review of the manuscript and their assistance. They are also grateful to several previous students and researchers whose collaborative research is discussed here. Finally, they are grateful to NSREC of Canada, the Canada Research Chair program and the CNRS of France for supporting this research.

C.23 Gallium Arsenide by Mike Brozel

The author is delighted to acknowledge the help given to him over many years by his colleagues and friends both at UMIST and in industry. Specifically, he wishes to thank R. Blunt, I.R. Grant, and R. H. Wallis for their careful and critical reading of this manuscript.

C.24 High-Temperature Electronic Materials: Silicon Carbide and Diamond*by Magnus Willander, Milan Friesel, Qamar-ul Wahab, Boris Straumal*

Magnus Willander and Milan Friesel would like to thank Dr. V. Narayan for checking the text, and Dr. A. Baranzahi for letting us use Figs. 24.2–24.3. Qamar-ul Wahab thanks Mr. Amir Karim for all his support.

C.25 Amorphous Semiconductors: Structure, Optical, and Electrical Properties*by Kazuo Morigaki, Chisato Ogiwara*

We wish to thank M. Ichihara, K. Suzuki and M. Yamaguchi, Institute for Solid State Physics, University of Tokyo, for providing us with their unpublished materials (Fig. 25.7a,b). Stimulating and helpful discussions were held with Prof. S. Kugler during the stay of one of us (K.M.) at the Budapest University of Technology and Economics, for which K.M. is grateful.

C.28 Dielectric Materials for Microelectronics*by Robert M. Wallace*

RMW gratefully acknowledges the many discussions and hard work of his colleagues and students engaged in gate-stack research. This work is supported in part by the Texas Advanced Technology Program and the Semiconductor Research Corporation.

C.29 Thin Films*by Robert D. Gould[†]*

The author wishes to acknowledge the general support and encouragement of Prof. C. A. Hogarth, Department of Physics, Brunel University, and of Prof. W. Fuller, Department of Physics, Keele University. Particular thanks are also due to Prof. E. W. Williams, Electronic Engineering Group, Keele University for collaborative work and permission to reproduce Figs. 29.6 and 29.7.

D.33 Electron Transport Within the III–V Nitride Semiconductors, GaN, AlN, and InN: A Monte Carlo Analysis*by Brian E. Foutz, Stephen K. O'Leary, Michael Shur, Lester F. Eastman*

Financial support from the Office of Naval Research and the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. The use of equipment granted from the Canada Foundation for Innovation, and equipment loaned from the Canadian Microelectronics Corporation, is also acknowledged.

D.36 II–VI Narrow-Bandgap Semiconductors for Optoelectronics*by Ian M. Baker*

The author wishes to express his gratitude to Mike Kinch of DRS Technologies, Kadri Vural and Jose Arias of Rockwell/Boeing and Marion Reine and coworkers at BAE SYSTEMS, Lexington for supplying material for this chapter and valuable advice. Also the advice and support from my technical colleagues, particularly: Peter Capper, Chris Maxey, Chris Jones and Les Hipwood, and my management here at SELEX Infrared, particularly Graham Hall. Thanks also to my wife, Lesley, for help with the English.

D.37 Optoelectronic Devices and Materials*by Stephen Sweeney, Alfred Adams*

It is a pleasure to acknowledge the many people with whom the authors have worked with over the years. In particular, we would like to thank the staff and students, past and present, at the University of Surrey for their wide-ranging contributions to this work. We would also like to thank the editor, Safa Kasap, for his support and encouragement in preparing this chapter. On a personal note, SJS would like to thank his wife for her support whilst writing this chapter.

D.44 Optical Nonlinearity in Photonic Glasses*by Keiji Tanaka*

The author would like to thank his students, K. Sugawara and N. Minamikawa, for preparing illustrations and giving comments.

E.48 Photoconductors for X-Ray Image Detectors*by M. Zahangir Kabir, Safa Kasap, John Rowlands*

We would like to thank Dr. Randy Luhta for useful discussions. We acknowledge financial support from NSERC.

E.52 High-Temperature Superconductors*by Rainer Wesche*

I wish to thank P. Bruzzone for his support and encouragement. The careful reviewing and many constructive suggestions on the manuscript by J.F. Crawford are especially appreciated.

E.53 Molecular Electronics*by Michael Petty*

The author would like to thank the staff and students of, and visitors to, the Durham Centre for Molecular and Nanoscale Electronics, who have contributed to some of work described in this chapter. Particular thanks are due

to Fraser Stoddart for useful discussions on the rotaxane switching devices and for providing Fig. 53.17a. Thanks are also due to Karl Coleman for providing the drawings of carbon nanotubes, Fig. 53.7.

E.54 Organic Materials for Chemical Sensing

by Asim Kumar Ray

The authors are grateful to the Engineering & Physical Research Council (UK) and the European Union for financial support of their work over the last 20 years.

Gratitude is also due to Dr. S. C. Thrope of the Health Safety Executive (UK), Dr. M. Hofton of TQ Environmental Plc., Prof. K. C. Thompson of ALControl and Dr. F. A. Grunfeld of NIMA Technology for their collaboration and support. Prof. M. J. Cook of the University of East Anglia remains a generous provider of novel phthalocyanine compounds for sensing experiments over the years. Above all, the contributions from Drs. R. Capan, O. Omar and A. V. Nabok are acknowledged with gratitude.

About the Authors



Martin Abkowitz

Webster, NY, USA
mabkowitz@mailaps.org,
abkowitz@chem.chem.rochester.edu

Chapter D.39

Martin A. Abkowitz received his Ph.D. in Physics from Syracuse University in 1964. During the period 1964–65, Abkowitz was Andrew Mellon Postdoctoral Fellow in Physics at the University of Pittsburgh. In 1965, Abkowitz joined the Webster Research Center (now the Wilson Center for Research and Technology) of Xerox Corporation where he was a Principal Scientist until retirement in 1999. Abkowitz is currently a Visiting Scientist at the University of Rochester. He is a fellow of the American Physical Society. He has 174 publications including 35 US patents. Abkowitz has made over 250 contributed and invited presentations at international conferences.

Sadao Adachi

Gunma University
 Department of Electronic Engineering,
 Faculty of Engineering
 Gunma, Japan
adachi@el.gunma-u.ac.jp



Chapter D.31

Sadao Adachi received his Ph.D. from Osaka University and is Professor of Electrical Engineering at Gunma University. From 1980 to 1988 he was with NTT Electrical Communication Laboratories, Japan. He has published and presented over 200 technical papers and 20 textbooks on semiconductor physics and technology. His current research interests include physical properties of semiconductors and new functional materials.

Alfred Adams

University of Surrey
 Advanced Technology Institute
 Surrey, UK
alf.adams@surrey.ac.uk



Chapter D.37

Alfred Adams studied at Leicester University, UK, and in 1964 embarked on postdoctoral research at the University of Karlsruhe, Germany. His work on III–V semiconductors started in 1967 at the University of Surrey where he is now a Distinguished Professor. He received the Duddell medal from the Institute of Physics in 1995 for proposing the use of strained quantum wells in III–V lasers and was elected Fellow of the Royal Society in 1996.



Guy J. Adriaenssens

University of Leuven
 Laboratorium voor Halfgeleiderfysica
 Leuven, Belgium
guy.adri@fys.kuleuven.ac.be

Chapter A.7

Guy Adriaenssens received his Ph.D. from the University of Washington, Seattle, in 1971. After a postdoctoral stay at the University of the Saarland, Germany, he joined the University of Leuven, Belgium, in 1973. His current research interests center on transport properties and the electronic density of states of amorphous semiconductors and chalcogenide glasses.



Wilfried von Ammon

Siltronic AG
 Research and Development
 Burghausen, Germany
wilfried.ammon@siltronic.com

Chapter A.5

Dr. von Ammon studied physics at the Technical University of Munich and the University of Regensburg. In 1981, he received his Ph.D. in physics from the University of Regensburg. In 1982, he joined Wacker Siltronic and has been working in research and development since then. His main focus was on silicon crystal growth (floating zone and Czochralski method), silicon-layer growth by chemical vapor deposition, neutron transmutation doping, thermal treatment of silicon and behavior of defects in silicon. In 1994, he managed a research project in collaboration with Sematech, USA, to assess 400 mm technology for the decision on the next-generation wafer. During recent years he has primarily been working on the development and improvement of 300 mm crystal-growth processes and on defect engineering of the silicon bulk.

Peter Ashburn

University of Southampton
School of Electronics
and Computer Science
Southampton, UK
pa@ecs.soton.ac.uk



Chapter C.22

Peter Ashburn received a Ph.D. degree in electrical and electronic engineering in 1974 from the University of Leeds and then joined the Philips Research Laboratories. In 1978 he joined the University of Southampton and is currently a Professor of Microelectronics. His research interests include SiGe heterojunction bipolar transistors (HBTs), ultimate CMOS and carbon nanotubes. He has published over 200 papers in the scientific literature and has authored two books on bipolar transistors.

Mark Auslender

Ben-Gurion University
of the Negev Beer Sheva
Department of Electrical
and Computer Engineering
Beer Sheva, Israel
marka@ee.bgu.ac.il



Chapter C.21

Mark Auslender received a Ph.D in solid-state theory in 1977. He was a senior researcher at Institute of Metal Physics, Ural Branch of the Academy of Sciences of the USSR. In 1991 he joined the Microelectronics Laboratory at the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel as a grade A researcher. His present interests focus on micro- and nanometer-scale diffraction gratings, optical coherence in regular and disordered media.

Darren M. Bagnall

University of Southampton
School of Electronics
and Computer Science
Southampton, UK
dmb@ecs.soton.ac.uk

Chapter C.22

Dr. Darren Bagnall is a senior lecturer in the School of Electronics and Computer Science at Southampton University. His research has included pioneering work on molecular-beam epitaxy (MBE) and chemical vapor deposition (CVD) for the development of new nano and quantum devices based on ZnO, silicon and silicon germanium. He has published over 50 refereed papers.

Ian M. Baker

SELEX Sensors
and Airborne Systems Infrared Ltd.
Southampton, Hampshire, UK
ian.m.baker@selex-sas.com

Chapter D.36

Ian Baker received a Ph.D. in solid-state imaging devices at Southampton University in 1973. His career in solid-state imaging has spanned the development of visible imaging charge-coupled devices (CCDs) within Philips and second- and third-generation infrared detectors at various companies. He has published over 75 papers and has over 30 patents in the field of electro-optics. His current interest is the development of advanced third-generation infrared detectors, including 3D imaging.

Sergei Baranovskii

Philipps University Marburg
Department of Physics
Marburg, Germany
baranovs@staff.uni-marburg.de



Chapter A.9

Professor Sergei Baranovskii received his Ph.D. (1981) in theoretical physics from the Ioffe Physical-Technical Institute of the Russian Academy of Sciences in St. Petersburg, where he worked as a senior researcher until 1990. Since 1990 he has been working at the Philipps University Marburg, Germany, where he got a Habilitation in Theoretical Physics in 1995. His research interests are devoted to charge transport and optical properties of organic and inorganic disordered solids.

Mark Baxendale

Queen Mary, University of London
Department of Physics
London, UK
m.baxendale@qmul.ac.uk



Chapter E.50

Mark Baxendale is a Reader in Nanotechnology at Queen Mary, University of London since 2002. The focus of his research is the physics and applications of carbon nanotubes. These applications include molecular quantum electronic devices, probes for scanning probe microscopy and single-molecule detection.

**Mohammed L. Benkhedir**

Chapter A.7

University of Leuven
Laboratorium voor Halfgeleiderfysica
Leuven, Belgium
MohammedLoufti.Benkhedir
@fys.kuleuven.ac.be

Mohammed Benkhedir received an M.S. in Physics from the University of Annaba, Algeria, and holds a Lectureship at the Centre Universitaire de Tebessa, Algeria. He is currently studying the electronic properties and density of states in amorphous selenium by means of photoconductivity techniques at the University of Leuven, Belgium.

**Monica Brinza**

Chapter A.7

University of Leuven
Laboratorium voor Halfgeleiderfysica
Leuven, Belgium
monica.brinza@fys.kuleuven.ac.be

Monica Brinza received an M.S. in Physics from the University of Bucharest, Romania and a Ph.D. from the University of Leuven, Belgium. Her current research interests focus on the transport properties of materials for photovoltaic applications and the electronic density of states of amorphous semiconductors and chalcogenide glasses.

Paul D. Brown

Chapter B.17

University of Nottingham
School of Mechanical, Materials
and Manufacturing Engineering
Nottingham, UK
paul.brown@nottingham.ac.uk



Paul D. Brown received a Ph.D. in Applied Physics from the University of Durham in 1989. He is presently Reader in Materials Characterisation at the University of Nottingham. His present research interests are concerned with the interrelationship between the structure, property and processing of structural, functional and biomedical materials, assessed using a broad range of characterisation techniques, with emphasis on novel variants of electron microscopy.

Mike Brozel

Chapter C.23

University of Glasgow
Department of Physics and Astronomy
Glasgow, UK
mikebrozel@beeb.net



For most of his career, Mike Brozel has worked on the growth and assessment of semiconductors. Present investigations include instrumentation for the mapping of bulk GaAs and novel structures for silicon nuclear particle detectors. An author of over 100 papers, he was the co-author with G. E. Stillman of *Properties of Gallium Arsenide*, 3rd ed. and recently co-authored a chapter on GaAs in *Bulk Crystal Growth*.

**Lukasz Brzozowski**

Chapter D.45

University of Toronto
Sunnybrook and Women's Research
Institute, Imaging Research/ Department
of Medical Biophysics
Toronto, ON, Canada
lukbroz@sten.sunnybrook.utoronto.ca

Lukasz Brzozowski has extensive experience in optics, medical imaging, nanotechnology and product development. He is currently leading research on the development of a combined X-ray/magnetic resonance imaging (MRI) system, and on the application of photo-excitable quantum dots for the treatment of cardiovascular disease. Dr. Brzozowski was awarded the Governor General's Gold Medal for the best applied Ph.D. from the University of Toronto in the year 2003.

**Peter Capper**

Chapters B.12, B.14, B.15

SELEX Sensors
and Airborne Systems Infrared Ltd.
Materials Team Leader
Southampton, Hampshire, UK
pete.capper@selex-sas.com

Peter Capper has spent some 30 years in the infrared industry growing and characterising cadmium mercury telluride (CMT) and other tellurium-based materials. He holds a patent in the bulk growth of CMT, has authored/co-authored over 100 papers and given several invited talks at international crystal growth/infrared (IR) conferences. He has edited/co-edited five books in IR materials and devices and crystal growth.

Larry Comstock

San Jose State University
San Jose, CA, USA
Comstock@email.sjsu.edu



Chapter E.51

R. Lawrence Comstock, Ph.D, has over 28 years of experience in computer storage system development. He was an IBM engineering manager for 25 years and was Vice President, Advanced Recording Technology at Maxtor Corporation. He received his Ph.D. in electrical engineering from Stanford University, and is the author of several books and publications. Larry has been teaching courses to graduate students at institutions such as Stanford University since 1993. He received the IEEE Fellow Award for Developments in Magnetic Storage and Magnetic Microwave Devices.

Ray DeCorby

University of Alberta
Department of Electrical
and Computer Engineering
Edmonton, Alberta, Canada
rdecorby@trlabs.ca



Chapter D.43

Ray DeCorby received his M.Sc. degree in Electrical Engineering from the University of Saskatchewan in 1995, and his Ph.D. degree in Electrical and Computer Engineering from the University of Alberta in 1998. Currently, he is an Associate Professor of Electrical and Computer Engineering at the University of Alberta. His research interests include integrated optics on silicon platforms, photonic glasses and polymers, and integrated nonlinear optics.

M. Jamal Deen

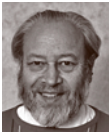


McMaster University
Department of Electrical and Computer
Engineering (CRL 226)
Hamilton, Canada
jamal@mcmaster.ca

Chapter B.20

M. Jamal Deen is Professor and Senior Canada Research Chair in Information Technology at McMaster University, Canada. His research work is documented in 14 invited book chapters, six awarded patents, and more than 300 peer-reviewed articles. His current research interests are in microelectronics and optoelectronics. Dr. Deen was a Fulbright–Laspau Scholar, an American Vacuum Scholar, won the Electrochemical Society (ECS) Callinan Award and the Distinguished Researcher Award from the Province of Ontario. He is a fellow of the IEEE, fellow of the EIC (Engineering Institute of Canada) and a fellow of the ECS.

Leonard Dissado



The University of Leicester
Department of Engineering
Leicester, UK
lad4@le.ac.uk

Chapter A.10

Professor Dissado was awarded a D.Sc. in 1990 by The University of London for his work on the theory of dielectric response and electrical breakdown phenomena. His research interests in dielectric response covers a wide range of materials from glasses and high voltage insulators, to bio-tissues. He is also co-author of the book *Electrical Degradation and Breakdown in Polymers*.

David Dunmur

University of Southampton
School of Chemistry
Southampton, UK
d.a.dunmur@soton.ac.uk



Chapter D.38

David Dunmur received his D.Phil. degree from the University of Oxford in 1965. His research over more than 30 years has been concerned with the physical properties of liquid crystals, specifically their dielectric, optical, electro-optical and elastic properties. He was founding editor of *Liquid Crystals Today*, and the 1999 recipient of the G. W. Gray Medal of the British Liquid Crystal Society. Professor Dunmur is an Emeritus Professor of Chemistry at the University of Sheffield, and currently a Visiting Professor at the University of Southampton.

Lester F. Eastman

Cornell University
Department of Electrical
and Computer Engineering
Ithaca, NY, USA
lfe2@cornell.edu



Chapter D.33

Lester F. Eastman is presently the John L. Given Foundation Chair Professor of Engineering at Cornell University. He joined the faculty of Electrical Engineering at Cornell in 1957, and also serves as a member of the graduate fields of Applied Physics and Materials Science. Since 1965 he has been doing research on compound semiconductor materials, high-speed devices and circuits, and has been active in organizing workshops and conferences on these subjects elsewhere since 1965 and at Cornell from 1967. He has supervised over 100 Ph.D. theses, over 50 M.S. theses, and over 50 postdoctoral studies. In his research group effort is underway on molecular-beam epitaxy, microwave transistors, high-speed semiconductor lasers, and fundamental phenomena in compound semiconductor quantum electron and optical devices. In 1991 he was awarded the GaAs Symposium Award and the Heinrich Welker medal. He was awarded the Alexander von Humboldt Senior Fellowship in 1994, and the Aldert van der Ziel Award in 1995. The IEEE honored him with their 1999 Graduate Teaching Award and Third Millennium Medal 2000. He was recently elected Fellow of the American Physical Society, 2001 and received the University of Michigan's William Gould Dow Lectureship award the same year. The biennial IEEE Cornell Conference on High-Performance Devices was renamed the IEEE Lester Eastman Conference in 2002. The Electron Devices Society has selected him for the 2002 recipient of the IEEE Electron Devices Society J. J. Ebers. Award. Distinguished Educator Award, 2003, for IEEE Microwave Theory and Technique Society.

Andy Edgar

Victoria University
School of Chemical
and Physical Sciences SCPS
Wellington, New Zealand
Andy.Edgar@vuw.ac.nz

Chapter D.40

Dr. Edgar is an Associate Professor in the School of Chemical and Physical Sciences at Victoria University, Wellington, New Zealand, and a Principal Investigator of the MacDiarmid Institute. His current research interests are storage phosphors for radiation imaging, and glass ceramics for optoelectronic applications.

Brian E. Foutz

Cadence Design Systems
Endicott, NY, USA
foutz@cadence.com

Chapter D.33

Brian E. Foutz is currently a Senior Member of Consulting Staff with Cadence Design Systems. Dr. Foutz's current research focuses on ASIC design-for-test strategies. Dr. Foutz is currently a member of the Institute of Electrical and Electronics Engineers.

Mark Fox

University of Sheffield
Department of Physics and Astronomy
Sheffield, UK
mark.fox@shef.ac.uk



Chapter D.42

Dr. Fox obtained his D.Phil. degree in physics from Oxford University in 1987. After post-doctoral work at AT&T Bell Laboratories, he was a Royal Society University Research Fellow at Oxford until 1998, when he took up his present post as a Reader in Physics at Sheffield University. His research interests include low-dimensional semiconductor structures, quantum optics and ultrafast laser spectroscopy.

Darrel Frear

RF and Power Packaging Technology
Development, Freescale Semiconductor
Tempe, AZ, USA
darrel.frear@freescale.com



Chapter E.55

Darrel Frear has an A.B. in Engineering Science from Dartmouth College and M.S. and Ph.D. degrees in Materials Science from the University of California, Berkeley. He is manager of Radio-frequency (RF) and Power Packaging Technology Development at Freescale Semiconductor. Previously, Darrel was at Sandia National Labs. Darrel's background includes materials research and development to enhance processing and reliability of electronic components.



Milan Friesel

Chapter C.24

Chalmers University of Technology
Department of Physics
Göteborg, Sweden
friesel@chalmers.se

Milan Friesel obtained his Ph.D. in Experimental High-Pressure Physics from Chalmers University of Technology in Göteborg in 1987. From 1988 to 1991 he was guest researcher at the Max Planck Institute for Metal Research in Stuttgart. Since 1994 he is Associate Professor in physics at the SIMS laboratory, Chalmers University of Technology. His research interests are superionic conductors, metals and alloys, semiconductors and characterization by the SIMS technique.



Jacek Gieraltowski

Chapter A.4

Université de Bretagne Occidentale
29285 Brest Cedex, France
Jacek.Gieraltowski@univ-brest.fr

Professor Gieraltowski holds a DSc. from Warsaw University (Poland) on the basis of research done in a French CNRS laboratory in Meudon, Bellevue (Paris, France). He has been involved in research on magnetic materials (ferrites, thin films and nanowires) for more than 30 years with numerous journal papers. After several years as a senior researcher in the Res. Lab. of Polfa in Warsaw associated with the Polish Academy of Science in Warsaw, he joined the department of Physics, in Brest (France) as an assistant professor working on the physics of ferrites and magnetic materials for microwave devices. Presently, he is full professor of Physics in Brest, working on fast switching magnetization problems for high density magnetic recording, giant magnetoimpedance sensors and the application of magnetism to biological and medical applications.

Yinyan Gong



Chapter D.35

Columbia University
Department of Applied Physics
and Applied Mathematics
New York, NY, USA
yg2002@columbia.edu

Yinyan Gong is graduate student of Columbia University. Her Current research interests are optical and electrical properties as well as microstructures of wide-band-gap materials. She received a Guanhua fellowship from Shanghai Jiao Tong University in 1997/1998.

Robert D. Gould[†]



Chapter C.29

Keele University
Thin Films Laboratory, Department of
Physics, School of Chemistry and Physics
Keele, Staffordshire, UK

Dr. Robert Gould has received two doctorates from Brunel University, a Ph.D. degree in 1973, and a D.Sc. degree in 2000 for his contribution to knowledge in the physics and technology of thin films. He spent four years lecturing in Physics at the University of Science, Malaysia, before joining Keele University in 1980, where he is currently a Senior Lecturer in Physics. He has published extensive review articles on electrical conduction in both phthalocyanines and cadmium compounds, and also on high-field conduction in nanostructures. He also has research interests in insulating nitride films prepared by sputtering. Dr. Robert Gould died suddenly on March 16, 2006 and will be greatly missed by the scientific community. *Editors.*



Shlomo Hava

Chapter C.21

Ben-Gurion University
of the Negev Beer Sheva
Department of Electrical
and Computer Engineering
Beer Sheva, Israel
hava@ee.bgu.ac.il

Shlomo Hava received a Ph.D. in Electrical Engineering from the University of Delaware in 1982. Since then he has worked at the Department of Electrical and Computer Engineering at Ben-Gurion University of the Negev, Israel. Currently he is a faculty professor and head of the Microelectronics Laboratory at this department. His present research interests include micro- and nanometer-scaled diffraction gratings in optical elements and light emitters, and the effects of gamma radiation and vacuum on optoelectronic devices.



Colin Humphreys

University of Cambridge
Department of Materials Science
and Metallurgy
Cambridge, UK
colin.Humphreys@msm.cam.ac.uk

Chapter .1

Colin Humphreys is the Goldsmiths' Professor of Materials Science at Cambridge University. He is the Director of the Cambridge Centre for Gallium Nitride and has published over 500 research papers, mainly on electron microscopy and analysis, semiconductors, superconductors and high-temperature alloys. He has received various national and international medals and awards for his research.

Stuart Irvine

University of Wales, Bangor
Department of Chemistry
Gwynedd, UK
sjc.irvine@bangor.ac.uk



Chapters B.14, E.46

Professor Irvine is a Chartered Physicist and received his B.Sc degree in Physics from Loughborough University of Technology, Ph.D. in Physical Metallurgy and Science of Materials, and D.Sc. in Physics, both from the University of Birmingham. His research covers various aspects of metalorganic chemical vapor deposition (MOCVD) of compound semiconductor and oxide materials for applications such as solar cells, detectors and displays. This research has led to the development of new optical in situ monitoring devices that are now manufactured and sold worldwide. Previous appointments include Rockwell Science Centre and the Royal Signals and Radar Establishment, Malvern where he carried out pioneering research in MOCVD of II–VI semiconductors.

Minoru Isshiki

Tohoku University
Institute of Multidisciplinary Research
for Advanced Materials
Sendai, Japan
ishiki@tagen.tohoku.ac.jp



Chapters B.16, D.34

Minoru Isshiki is a Professor at the Tohoku University in the Institute of Multidisciplinary Research for Advanced Materials. He earned his Ph. D. in materials science at Tohoku University in 1976. Dr. Isshiki is working on the purification of metals, single-crystal growth of bulk and thin-film compound semiconductors and impurity effects on their properties.



Robert Johanson

University of Saskatchewan
Department of Electrical Engineering
Saskatoon, SK, Canada
johanson@engr.usask.ca

Chapter A.2

Robert Johanson is currently an Associate Professor in the Department of Electrical Engineering at the University of Saskatchewan, Canada. He obtained his A.B. (honors) (1981), M.Sc. (1983) and Ph.D. (1991) from the University of Chicago, and subsequently worked as a postdoctoral research fellow and a research associate at the University of Saskatchewan. His research interests include electrical and optical properties of amorphous semiconductors, and glasses. Dr. Johanson has published numerous journal and conference papers in this field, and is a member of the American Physical Society and the IEEE.



Tim Joyce

University of Liverpool
Functional Materials Research Centre,
Department of Engineering
Liverpool, UK
t.joyce@liv.ac.uk

Chapter B.14

Tim Joyce has worked on semiconductor epitaxy since obtaining his B.Sc. in 1981 and was awarded a Ph.D. in Materials Science in 1991 at the University of Liverpool for research in metalorganic molecular-beam epitaxy (MOMBE). Current research interests include MBE of group III–N and group III–V–N semiconductors. He is currently chairman of the British Association of Crystal Growth.

M. Zahangir Kabir

Concordia University
Department of Electrical
and Computer Engineering
Montreal, Quebec, Canada
kabir@encs.concordia.ca



Chapter E.48

Zahangir Kabir received the B.Sc. and M.Sc. degrees in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Bangladesh, in 1994 and 1996, respectively. From 1996 to 2000 he was an Assistant Professor with the Electrical and Electronic Engineering department at BUET. He received his PhD from the University of Saskatchewan in 2005, and is currently Assistant Professor at Concordia University in Montreal. His research interests are in the areas of imaging sensors, disordered semiconductors, and electronic materials and devices.

Safa Kasap

University of Saskatchewan
Department of Electrical Engineering
Saskatoon, SK, Canada
safa.kasap@usask.ca



Chapters [A.2](#), [A.3](#), [B.19](#), [E.48](#)

Safa Kasap is currently a Professor and Canada Research Chair in Electronic Materials and Devices in the Electrical Engineering Department at the University of Saskatchewan, Canada. He obtained his B.Sc. (1976), M.Sc. (1978) and Ph.D. (1983) degrees from the Imperial College of Science, Technology and Medicine, University of London, specializing in amorphous semiconductors and chalcogenide glasses. In 1996 he was awarded the D.Sc. (Engineering) from London University for his research contributions to materials science in electrical engineering. He is a fellow of the Institution of Electrical Engineers, the Institute of Physics and the Institute of Materials. His research interests are in amorphous semiconductors, glasses for photonics, photoconductors, electrical, optical and thermal properties of materials, and related topics, with more than one hundred refereed journal papers in these areas. He is the author of *Optoelectronics and Photonics: Principles and Practices* and *Principles of Electronic Materials and Devices*, both textbooks are widely used by many major universities and professionals.

Alexander Kolobov



National Institute of Advanced Industrial
Science and Technology
Center for Applied
Near-Field Optics Research
Ibaraki, Japan
a.kolobov@aist.go.jp

Chapter [E.49](#)

Dr. A. V. Kolobov obtained his Ph.D. from the A. F. Ioffe Physico-Technical Institute at St. Petersburg, Russia. Since 1994 he has been a senior researcher at the National Institute of Advanced Industrial Science and Technology (AIST) in Japan. His current interests include, but are not limited to, amorphous semiconductors, photoinduced phenomena in solids, optical data storage and use of synchrotron radiation in materials science.

Cyril Koughia



University of Saskatchewan
Department of Electrical Engineering
Saskatoon, SK, Canada
kik486@mail.usask.ca

Chapters [A.2](#), [A.3](#)

Cyril Koughia received his Ph.D. from the A. F. Ioffe Physico-Technical Institute, St. Petersburg, Russia and is a Research Associate at the University of Saskatchewan, Canada. His current research interests include the physical properties of amorphous semiconductors and glasses used in photonics and electronics.

Igor L. Kuskovsky

Queens College,
City University of New York (CUNY)
Department of Physics
Flushing, NY, USA
igor_kuskovsky@qc.edu



Chapter [D.35](#)

Igor L. Kuskovsky is an Assistant Professor of Physics at Queens College, CUNY. Research interests include optical and magneto-optical properties of quantum dots and nanowires, applications of low dimensional systems for bio-imaging, and microstructural, electrical and electro-optical properties of wide-band-gap semiconductors. Previously, he has done extensive work on optical properties of II–VI wide-band-gap semiconductors.

Geoffrey Luckhurst

University of Southampton
School of Chemistry
Southampton, UK
g.r.luckhurst@soton.ac.uk



Chapter [D.38](#)

Geoffrey Luckhurst was awarded his Ph.D. degree in 1965 from the University of Cambridge having studied for his B.Sc. at the University of Hull. His research into liquid crystals started in Cambridge and continues at Southampton where he is the Emeritus Professor of Chemical Physics. His work is concerned with many aspects of this interdisciplinary field including the design of novel materials, their molecular orientational order, the molecular dynamics, their macroscopic order and dynamics using both experimental and theoretical techniques. He is co-founder of the international journal, *Liquid Crystals*, was the recipient of the G. W. Gray Medal of the British Liquid Crystal Society in 2002 and was elected President of the International Liquid Crystal Society of which he is an Honoured Member.

**Akihisa Matsuda**

Chapter C.26

Tokyo University of Science
Research Institute for Science
and Technology
Chiba, Japan
amatsuda@rs.noda.tus.ac.jp, matsuda@aist.go.jp

Dr. Akihisa. Matsuda has an M.S. degree from Waseda University, a Ph.D. from the Tokyo Institute of Technology, and is Chief Senior Researcher at ETL MITI Japan. He is also the Director of the Research Initiative for Thin-Film Silicon Solar-Cells AIST Japan.

**Naomi Matsuura**

Chapter D.41

Sunnybrook Health Sciences Centre
Department of Medical Biophysics,
Imaging Research
Toronto, ON, Canada
matsuura@sri.utoronto.ca

Naomi Matsuura received her Ph.D. from the University of Toronto in 2003 for the development of 2D and 3D periodic, three-dimensional nanoscale architectures using non-lithographic parallel patterning technologies. Dr. Matsuura is currently a Canadian Institutes of Health Research (CIHR) Strategic Training Fellow in the EIRR21st Program at Toronto's Sunnybrook Health Sciences Centre, and is investigating the development of novel nanostructures for cancer imaging and therapy.

Kazuo Morigaki

Chapter C.25

University of Tokyo
Tokyo, Japan
k.morigaki@yacht.ocn.ne.jp



Kazuo Morigaki received his Ph.D. in physics from Osaka University in 1959. Since then, he has joined Osaka University, Sony Corporation Research Laboratory, Centre d'Etudes Nucléaires de Saclay, University of Tokyo (Institute for Solid State Physics), Yamaguchi University, and Hiroshima Institute of Technology. He is now Professor emeritus at the University of Tokyo. His current area of interest are light-induced phenomena and the electronic states of defects in amorphous and microcrystalline silicon.

Hadis Morkoç

Chapter D.32

Virginia Commonwealth University
Department of Electrical
and Computer Engineering
Richmond, VA, USA
hmorkoc@vcu.edu



Professor Hadis Morkoç received his Ph.D. from Cornell University, and is currently the Founders Professor of Electrical Engineering and Physics at the Commonwealth Virginia University in Richmond, Virginia, where his research interest cover GaN and related III–nitride semiconductors, ZnO and perovskite oxides, and quantum wells and quantum-dot structures. Professor Morkoç has authored and coauthored a number of books on optoelectronic materials and devices, such as *Nitride Semiconductors and Devices*, *Advanced Semiconductor and Organic Nano-Techniques*, *Principles and Technology of MODFETS*. He has published some 1300 technical articles in electronic and optoelectronic materials and devices from fundamental materials issues to device applications. Professor Morkoç is a Fellow of IEEE, Fellow of American Association for the Advancement of Science and a Life Fellow of the American Physics Society.

**Winfried Mönch**

Chapter A.8

Universität Duisburg–Essen
Duisburg, Germany
w.moench@uni-duisburg.de

Winfried Mönch received a Dr. rer. nat. degree from the Universität Göttingen in 1961. He spent three years at the AEG-Forschungsinstitut before joining the RWTH Aachen. There he qualified as university lecturer in 1968 and became Associate Professor two years later. In 1974 he was appointed Professor at the Universität Duisburg and retired in 1999. He was Walter Schottky Visiting Professor at Stanford University in 1981 and received the E. W. Müller Award 1984 of the University of Wisconsin–Milwaukee. He has authored two monographs on semiconductor surfaces and interfaces.



Arokia Nathan

University of Waterloo
Department of Electrical
and Computer Engineering
Waterloo, Ontario, Canada
anathan@uwaterloo.ca

Chapter [E.47](#)

Arokia Nathan is a Professor in Electrical and Computer Engineering at the University of Waterloo, Canada. He held the DALSA/Natural Sciences and Engineering Research Council of Canada (NSERC) industrial research chair in sensor technology, and was a recipient of the 2001 NSERC E. W. R. Steacie Fellowship. He currently holds the Canada Research Chair in Nanoscale Elastic Circuits. His research interests lie in materials devices, and circuits pertinent to large-area electronics.

Gertrude F. Neumark

Columbia University
Department of Applied Physics and
Applied Mathematics
New York, NY, USA
gfn1@columbia.edu



Chapter [D.35](#)

Gertrude F. Neumark is Howe Professor of Materials Science. Current research interests are electrical and optical properties of wide bandgap semiconductors, mainly ZnSe and III–nitrides. She is Fellow of the American Physical Society and has a listing by the American Physical Society on the archival web site for Notable Women in Physics. She was awarded by the National Science Foundation (NSF) as Visiting Professor at Columbia University (1982/83), and is Panelist for the National Research Council (NRC) and NSF.

Stephen K. O’Leary

University of Regina
Faculty of Engineering
Regina, SK, Canada
stephen.oleary@uregina.ca



Chapters [A.3](#), [D.33](#)

Stephen K. O’Leary is currently an Associate Professor in the Faculty of Engineering at the University of Regina. Professor O’Leary’s research focuses on novel electronic materials and the electron devices that use such materials. He is currently a member of the Institute of Electrical and Electronics Engineers and the Materials Research Society.



Chisato Ogihara

Yamaguchi University
Department of Applied Science
Ube, Japan
ogihara@yamaguchi-u.ac.jp

Chapter [C.25](#)

Dr. C. Ogihara received his Ph.D. degree in physics from University of Tokyo in 1988. He joined the University of Strathclyde in 1989, Gifu University in 1991 and Yamaguchi University in 1993. His current research focuses on photoluminescence and light-induced creation of defects in hydrogenated amorphous silicon.



Fabien Pascal

Université Montpellier 2/CEM2–cc084
Centre d’Electronique
et de Microoptoelectronique de
Montpellier
Montpellier, France
pascal@cem2.univ-montp2.fr

Chapter [B.20](#)

Fabien Pascal is involved in a research group working in electronic devices (MOS, MODFET, MESFET, TBH Si–Ge, TBH III–V) by means of their proper background I – V / C – V characteristics and noise spectral analysis. He is also working on the technological qualification of contacts, semiconductor materials and carbon nanotubes based devices by the use of noise spectroscopy. He is author and co-author of more than 70 publications in international journals and conferences.

Michael Petty

University of Durham
Department School of Engineering
Durham, UK
m.c.petty@durham.ac.uk



Chapter [E.53](#)

Michael Petty obtained his Ph.D. from Imperial College, London for work on semiconducting II–VI thin films. His current research activities encompass nanoscale architectures of organic materials, e.g. build up by techniques such as Langmuir–Blodgett deposition, self-assembly, and thermal evaporation. He has a special interest in the application of such thin layers to electronic and optoelectronic devices. He is currently Co-Director of the Durham Centre for Molecular and Nanoscale Electronics.

Asim Kumar Ray

Queen Mary, University of London
Department of Materials
London, UK
a.k.ray@qmul.ac.uk



Chapter E.54

Asim K. Ray B.Sc., M.Sc., Ph.D., D.Sc, FRSA holds the Chair in Functional Materials at Queen Mary, University of London. His research interests lie in thin-film technologies for fabrication of nanostructures and chemical and biosensors. He is a fellow of the Institution of Electrical Engineers, UK and the Institute of Physics (UK). He is a college member of the Engineering and Physical Science Research Council, UK and an honorary editor of the Proc. IEE – Circuits, Devices, and Systems.

John Rowlands

University of Toronto
Department of Medical Biophysics
Sunnybrook and Women's College
Health Sciences Centre
Toronto, Canada
john.rowlands@sri.utoronto.ca

Chapter E.48

John Rowlands is a Professor of Medical Biophysics at the University of Toronto. His research laboratories devoted to the study of X-ray imaging and positron emission tomography are located in the Research Institute of Sunnybrook and Women's Health Sciences Centre. His research interests are the fundamental limitations of image quality in radiation imaging.

Oleg Rubel

Philipps University Marburg
Department of Physics
and Material Sciences Center
Marburg, Germany
oleg.rubel@physik.uni-marburg.de

Chapter A.9

Dr. Oleg Rubel received his Ph.D. in material sciences from the Zaporozhye National Technical University in Ukraine in 2001. Since 2003 he has been working as a member of the Central Technological Laboratory and of the Semiconductor Theory Group at the Philipps University Marburg. His current research activities include theoretical aspects of transport and optical properties of crystalline, amorphous and organic semiconductors where carrier localization plays a dominant role, as well as atomic-scale characterization techniques for semiconductor heterostructures.

Harry Ruda

University of Toronto
Materials Science and Engineering,
Electrical and Computer Engineering
Toronto, Canada
ruda@ecf.utoronto.ca



Chapters A.2, A.3, D.41

Professor Harry Ruda received his Ph.D. from MIT in 1982 in Materials Physics, and worked as an IBM postdoctoral fellow from 1982 to 1984, developing one of first theories for electron transport in selectively doped heterostructures. From 1984 to 1989 he was a senior scientist at 3M Corporation, developed some of first models for electronic transport and optical properties of wide-band-gap II–VI semiconductors. Currently he is a full professor and chair in nanotechnology at the University of Toronto, and Director of the Centre for Nanotechnology. He has published over 170 papers in international refereed journals and holds 14 patents. His current research focuses on the fabrication and modeling of quantum functional nanostructures with applications in the fields of nanoelectronics and nanophotonics.

Edward Sargent

University of Toronto
Department of Electrical
and Computer Engineering
Toronto, Canada
ted.sargent@utoronto.ca



Chapter D.45

Ted Sargent is 2004-5 Visiting Professor, Nanotechnology and Photonics, in the Microphotonics Center at MIT. In 2003 he was named one of the world's top young innovators by MIT's Technology Review. In 2002 he was honoured by the Canadian Institute for Advanced Research as one of Canada's top 20 researchers under the age of 40. In his research he innovates devices for networking, energy, and medicine by applying colloidal quantum dots, photonic crystals, and other advanced nanomaterials.



Peyman Servati

Chapter E.47

Ignis Innovation Inc.
Waterloo, Ontario, Canada
pservati@uwaterloo.ca

Peyman Servati received the B.Sc. degree from the University of Tehran in 1998, and his M.A.Sc. and Ph.D. degrees from the University of Waterloo, Ontario, Canada, in 2000 and 2004, respectively. His present research interest lies in disordered organic and silicon thin-film transistors (TFTs) and nanostructured materials. He was a recipient of the 2005 NSERC Doctoral Prize, NSERC Postgraduate Scholarship, and the Bronze Medal in the XXV I.Ph.O. competitions held in Beijing, China, 1994.



Derek Shaw

Chapter A.6

Hull University
Hull, UK
DerekShaw1@compuserve.com

Dr. Derek Shaw received his B.Sc. in physics from the University of Manchester in 1950 and his Ph.D. in meteorology from Imperial College London in 1955. After spending several years in industry working on thermionic emitters and photoconductors he joined the Physics Department of Hull University in 1963. There he initiated research in semiconductor diffusion with specific interest in Si, Ge, AlSb, GaAs, GaSb, CdTe, CdS, HgCdTe and PbSnTe. In retirement his interest in HgCdTe continues in collaboration with SELEX Sensors & Airborne Systems.

Fumio Shimura

Chapter B.13

Shizuoka Institute of Science and Technology
Department of Materials and Life Science
Fukuroi, Shizuoka, Japan
shimura@ms.sist.ac.jp



Dr. Shimura has been engaged in both fundamental and practical science and engineering related to semiconductor crystal technology and semiconductor device processing, and is the author of *Semiconductor Silicon Crystal Technology*. Fumio Shimura received his Ph.D. degree in Applied Physics from Nagoya University, Japan in 1982. He is currently a professor at the Shizuoka Institute of Science and Technology and Adjunct Professor, North Carolina State University.

Michael Shur

Chapter D.33

Rensselaer Polytechnic Institute
Department of Electrical, Computer,
and Systems Engineering
Troy, NY, USA
shurm@rpi.edu



Dr. Shur is Patricia and Sheldon Roberts Professor, Professor of Physics, and Director of Broadband Center at RPI. He is fellow of the IEEE, APS, ECS, World Innovation Foundation, and a winner of van der Ziel Award and of several best conference paper awards. He holds over 30 patents on solid-state devices and was listed by the ISA as a highly quoted researcher.



Jai Singh

Chapter A.3

Charles Darwin University
School of Engineering and Logistics,
Faculty of Technology, B-41
Darwin, NT, Australia
jai.singh@cdu.edu.au

Jai Singh is Professor of Physics at the Charles Darwin University, Darwin, Australia and Fellow of the Australian Institute of Physics. He has held several visiting fellowships, including the Max Planck Institute (Germany), Visiting Professor (USA), JSPS Fellow (Japan), and Guest Professor (Denmark). His research interests are in condensed-matter theory, covering areas of excitonic processes in crystalline and amorphous semiconductors, nanostructures, and designing of solar cells. He has written two books, edited one book and four conference proceedings, and published 140 research papers.



Tim Smeeton

Chapter .1

Sharp Laboratories of Europe
Oxford, UK
tim.smeeton@sharp.co.uk

Tim Smeeton received his Ph.D. from the University of Cambridge in 2005 following research into the nanostructures of InGa_N quantum wells using transmission electron microscopy and X-ray scattering techniques. He is now with Sharp Laboratories of Europe, where he is researching group III–nitride quantum-dot structures and devices.

Boris Straumal

Russian Academy of Sciences
Institute of Solid State Physics
Chernogolovka, Russia
straumal@issp.ac.ru



Chapter C.24

Professor Straumal is Head of the Laboratory for Interfaces in Metals at the Institute for Solid-State Physics of the Russian Academy of Sciences at Chernogolovka, Russia, and a Full Professor at the Moscow State Institute of Steel and Alloys (Technological University) in the Faculty of Physical Chemistry. His research interests focus on grain boundaries, phase transformations, diffusion, thermodynamics, coating technologies, properties of coatings, crystal growth, and nanomaterials.

Stephen Sweeney

University of Surrey
Advanced Technology Institute
Guildford, Surrey, UK
s.sweeney@surrey.ac.uk



Chapter D.37

Stephen John Sweeney is a Senior Lecturer (Associate Professor) in Semiconductor Laser Physics at the Advanced Technology Institute, University of Surrey, UK. He holds a B.Sc. (hons.) in Physics and a Ph.D. in Semiconductor Laser Physics. His research interests are in low-dimensional semiconductors for use in lasers and other photonic devices with applications including communications, bio-sensing and solid-state lighting.

David Sykes

Loughborough Surface Analysis Ltd.
Loughborough, UK
d.e.sykes@lsaltd.co.uk



Chapter B.18

David Sykes graduated from the University of Warwick in 1970 with a B.Sc. in Physics and obtained his Ph.D. from the University of Aston in 1975. Following postdoctoral positions at the Universities of Birmingham and York he moved to Loughborough University in 1978 where he remained until setting up LSA Ltd. in 1997. He is Director of Loughborough Surface Analysis Ltd. in Loughborough, UK which provides responsive, confidential, professional contract surface analysis service to industry and universities. Presently he is Chairman of the UK Surface Analysis Forum (UKSAF), Vice Chairman of the British Vacuum Council (BVC), Chairman of the ISO Technical Committee 201 Subcommittee on Data Management and Treatment in Surface Chemical Analysis (ISO TC201SC3), and Recording Secretary of the International Union for Vacuum Science, Technique and Application (IUVSTA).

Keiji Tanaka

Hokkaido University
Department of Applied Physics,
Graduate School of Engineering
Sapporo, Japan
keiji@eng.hokudai.ac.jp



Chapter D.44

After graduating from a master course at Hokkaido University in 1972, Professor Keiji Tanaka worked on photoreceptors at Canon Co. Ltd. for two years. Then, coming back to the university, and he was promoted to a professor in 1991. He is also head of the research group on Photoelectronics in Disordered Materials of the Japan Society of Applied Physics. He is interested in the optoelectronic properties of amorphous materials, specifically chalcogenide glasses, and applications. He received the first Ovshinsky Award in 2001.

Charbel Tannous

Université de Bretagne Occidentale
LMB, CNRS FRE 2697
Brest Cedex, France
tannous@univ-brest.fr



Chapter A.4

Charbel Tannous holds a D.Sc. from Joseph Fourier University (Grenoble, France) and a Ph.D. from the University of Sherbrooke (Quebec, Canada). After being a Postdoctoral Fellow at Cornell University (Ithaca, USA), he joined the Department of Engineering Physics, Montreal, where he worked on $1/f$ noise in fractal circuits and microelectronic device simulation. Later he joined the AGT research and development department in Calgary as a Senior Researcher working on wireless communications and nonlinear signal processing. After being at TRLabs and simultaneously Associate Professor of Electrical Engineering at the University of Saskatchewan, he joined the Université de Bretagne Occidentale in Brest, France, as a Full Professor of Physics. Presently he is working on fast switching magnetization problems for high-density magnetic recording and giant magnetoresistance sensors.

Ali Teke

Balikesir University
Department of Physics, Faculty of Art
and Science
Balikesir, Turkey
ateke@balikesir.edu.tr



Chapter D.32

Ali Teke is currently an Associate Professor in the Physics Department at the Balikesir University, Turkey. He received his B.Sc. (1992) and Ph.D. (1997) degrees in the Physics Departments of the Middle East Technical University, Turkey and Essex University, UK, respectively. His current research activities include the investigation of the electronic and optical properties of wide-band-gap semiconductors such as GaN and ZnO and exploring new devices for electronic and optoelectronic applications. He has published more than 30 journal and conference papers and chapters in books.

Junji Tominaga

National Institute of Advanced Industrial
Science and Technology, AIST
Center for Applied Near-Field Optics
Research, CAN-FOR
Tsukuba, Japan
j-tomonaga@aist.go.jp



Chapter E.49

Junji Tominaga received his Ph.D. in Materials Science from Cranfield Institute of Technology, UK, in 1991. After completing his Ph.D., he was a senior member of research staff at TDK Corporation and developed optical phase-change materials for rewritable CD and DVD. He joined AIST in 1997. Now, he is Director of the Center for Applied Near-Field Optics Research, CAN-FOR, Japan. His current research interests are high-density optical data storage and nanophotonics in nanotechnology.

Dan Tonchev

University of Saskatchewan
Department of Electrical Engineering
Saskatoon, SK, Canada
dan.tonchev@usask.ca



Chapter B.19

Dan Tonchev is currently a Research Associate in Electronic and Photonic Materials in the Electrical Engineering Department at the University of Saskatchewan, Canada. He is also a Professor in Chemical Technology at the Plovdiv University in Bulgaria. He obtained his M.Sc. (1981) from the Moscow State University, and Ph.D. (1990) in materials engineering from the University of Chemical Technology and Metallurgy, Sofia, Bulgaria. His research interests are in the preparation and characterization of glasses and polymers. He has published more than 40 journal and conference papers and chapters in books, and has 14 patents (US and European). He is a member of the American Chemical Society.

Harry L. Tuller

Massachusetts Institute of Technology
Department of Materials Science
and Engineering, Crystal Physics
and Electroceramics Laboratory
Cambridge, MA, USA
tuller@mit.edu



Chapter A.11

Harry Tuller received his Eng.Sc.D. from Columbia University in 1973 and joined the MIT faculty in 1975. Current research focuses on functional materials/MEMS integration and solid-state ionic-device optimization. He is Fellow of the American Ceramic Society, a von Humboldt awardee and recipient of *docteur honoris causa* of the Université de Provence, Marseille, France. He co-founded Boston Microsystems and is a pioneer in micromachined SiC-based sensor arrays.

Qamar-ul Wahab

Linköping University
Department of Physics, Chemistry,
and Biology (IFM)
Linköping, Sweden
quw@ifm.liu.se



Chapter C.24

Qamar-ul Wahab has a Master and M.Phil. degree in Physics from Qaid-i-Azam University, Islamabad, Pakistan. He earned his PhD in 1994 from Linköping University, Sweden in the area of silicon carbide materials and devices. Currently he is working as an Associate Professor at IFM, Linköping University and Senior Researcher at Swedish Defence Research (FOI). He worked with ABB Corp. Res. from 1995–1998 in the SiC power-device program designing and fabricating 3 kV Schottky barrier diode in 4H-SiC. Current research interests are in modelling and simulations of microwave power transistors and amplifiers.

**Robert M. Wallace**

Chapter C.28

University of Texas at Dallas
Department of Electrical Engineering
Richardson, TX, USA
rmwallace@utdallas.edu

Robert M. Wallace earned his Ph.D. in Physics from the University of Pittsburgh in 1988. In 1990, Wallace joined Texas Instruments and was appointed (1997) to manage materials research on advanced device concepts and the associated integration issues. In 2003, he joined the University of Texas at Dallas as a Professor of Electrical Engineering and Physics. He has over 85 publications and 65 US and international patents.

**Jifeng Wang**

Chapters B.16, D.34

Tohoku University
Institute of Multidisciplinary Research
for Advanced Materials
Sendai, Japan
wang@tagen.tohoku.ac.jp

Jifeng Wang received his Ph.D. degree from the Chinese Academy of Sciences in 1992. After a postdoctoral fellowship at Nagaoka University of Technology he became an Assistant Professor at the same University. From 1995 until now he has been working at Tohoku University. His current research focuses on the growth of II–VI compound semiconductors and environmental semiconductors.

David S. Weiss

NexPress Solutions, Inc.
Rochester, NY, USA
David_Weiss@Nexpress.com



Chapter D.39

David S. Weiss is a Scientist Fellow at NexPress Solutions, Inc. (a Kodak Company), in Rochester, New York. He received his Ph.D. in chemistry from Columbia University, New York in 1969. His research interests focus on electrophotographic technologies with emphasis on organic photoreceptors. He holds 16 U.S. patents and is author on over 70 publications. He is co-author of *Organic Photoreceptors for Imaging Systems*.

Rainer Wesche

Swiss Federal Institute of Technology
Centre de Recherches
en Physique des Plasmas
Lausanne, Switzerland
rainer.wesche@psi.ch



Chapter E.52

Rainer Wesche studied physics at the University of Constance, Germany. After completing his diploma in 1984, he was assistant from 1985 to 1989 (Ph.D. in 1988). From 1989 to 1993 he was a research scientist at the Paul Scherrer Institute in Switzerland, where he led an experimental study of high-current applications of high- T_c superconductors funded by the Swiss National Science Foundation. Since 1994 he has been a research scientist at the Swiss Federal Institute of Technology Lausanne. His present research is in the field of applied superconductivity.

**Roger Whatmore**

Chapter C.27

Tyndall National Institute
Lee Maltings, Cork, Ireland
roger.whatmore@tyndall.ie

Roger Whatmore is Professor of Engineering Nanotechnology, researching the use of ferroelectrics in microsystems and nanotechnology. He has published over 200 papers and 30 patents in the field. He is a Nelson Gold Medal and Griffith Medal winner, a fellow of the Royal Academy of Engineering and a fellow of the Institute of Materials, Minerals and Mining.

**Neil White**

Chapter C.30

University of Southampton
School of Electronics
and Computer Science
Highfield, Southampton, UK
nmw@ecs.soton.ac.uk

Neil White is Professor of Intelligent Sensor Systems in the School of Electronics and Computer Science at the University of Southampton, UK. He is a fellow of the Institution of Electrical Engineers (IEE) and the Institute of Physics (IOP) as well as a Senior Member of the IEEE. Professor White has published extensively in the field of sensor technology and novel sensing materials.

Magnus Willander

University of Gothenburg
Department of Physics
Göteborg, Sweden
mwil@fy.chalmers.se



Chapter C.24

Magnus Willander is chaired professor in physics at Gothenburg University in Sweden. He did pioneering works on Si/SiGe heterojunction bipolar transistors (HBTs), polymer transistors, and laser transistors. His interest in soft materials in the last five to six years led to the demonstration of the water transistor and its applications. Professor Willander combines experimental and theoretical research.

Jan Willekens

University of Leuven
Laboratorium voor Halfgeleiderfysica
Leuven, Belgium
jan.willekens@kuleuven.ac.be



Chapter A.7

Jan Willekens studied Physics at the University of Leuven, Belgium, and is currently a Physics Tutor at the Faculty of Engineering of the same University. He is using steady-state and transient photoconductivity techniques to study photogeneration and electronic transport in disordered semiconductor structures with special emphasis on polymeric blends for photovoltaic applications.

Detailed Contents

List of Abbreviations	XXIX
------------------------------------	------

1 Perspectives on Electronic and Optoelectronic Materials	
<i>Tim Smeeton, Colin Humphreys</i>	3
1.1 The Early Years	4
1.2 The Silicon Age	4
1.2.1 The Transistor and Early Semiconductor Materials Development	4
1.2.2 The Integrated Circuit	6
1.3 The Compound Semiconductors	8
1.3.1 High Speed Electronics	9
1.3.2 Light Emitting Devices	10
1.3.3 The III-Nitrides	12
1.4 From Faraday to Today	14
References	14

Part A Fundamental Properties

2 Electrical Conduction in Metals and Semiconductors	
<i>Safa Kasap, Cyril Koughia, Harry Ruda, Robert Johanson</i>	19
2.1 Fundamentals: Drift Velocity, Mobility and Conductivity	20
2.2 Matthiessen's Rule	22
2.3 Resistivity of Metals	23
2.3.1 General Characteristics	23
2.3.2 Fermi Electrons	25
2.4 Solid Solutions and Nordheim's Rule	26
2.5 Carrier Scattering in Semiconductors	28
2.6 The Boltzmann Transport Equation	29
2.7 Resistivity of Thin Polycrystalline Films	30
2.8 Inhomogeneous Media. Effective Media Approximation	32
2.9 The Hall Effect	35
2.10 High Electric Field Transport	37
2.11 Avalanche	38
2.12 Two-Dimensional Electron Gas	39
2.13 One Dimensional Conductance	41
2.14 The Quantum Hall Effect	42
References	44

3	Optical Properties of Electronic Materials: Fundamentals and Characterization	
	<i>Safa Kasap, Cyril Koughia, Jai Singh, Harry Ruda, Stephen K. O'Leary</i>	47
3.1	Optical Constants	47
3.1.1	Refractive Index and Extinction Coefficient	47
3.1.2	Kramers–Kronig Relations	49
3.2	Refractive Index	50
3.2.1	Cauchy Dispersion Equation	50
3.2.2	Sellmeier Dispersion Equation	51
3.2.3	Gladstone–Dale Formula	51
3.2.4	Wemple–Di Dominico Dispersion Relation	52
3.2.5	Group Index (N)	53
3.3	Optical Absorption	53
3.3.1	Lattice or Reststrahlen Absorption and Infrared Reflection	54
3.3.2	Free Carrier Absorption (FCA)	55
3.3.3	Band-to-Band or Fundamental Absorption	57
3.3.4	Exciton Absorption	63
3.3.5	Impurity Absorption	66
3.3.6	Effects of External Fields	69
3.4	Thin Film Optics	70
3.4.1	Swanepoel's Analysis of Optical Transmission Spectra	71
3.4.2	Ellipsometry	72
3.5	Optical Materials	74
3.5.1	Abbe Number or Constringence	74
3.5.2	Optical Materials	74
3.5.3	Optical Glasses	76
	References	76
4	Magnetic Properties of Electronic Materials	
	<i>Charbel Tannous, Jacek Gieraltowski</i>	79
4.1	Traditional Magnetism	81
4.1.1	Fundamental Magnetic Quantities	81
4.1.2	The Hysteresis Loop	83
4.1.3	Intrinsic Magnetic Properties	87
4.1.4	Traditional Types of Magnetism and Classes of Magnetic Materials	90
4.2	Unconventional Magnetism	93
4.2.1	Conventional and Unconventional Types of Exchange and Coupling in Magnetic Materials	93
4.2.2	Engineering and Growth of Thin Magnetic Films	94
4.2.3	Electronic Properties: Localized, Free and Itinerant Magnetism and Spin–Polarised Band Structure	95
4.2.4	Prospects for Spintronics and Quantum Information Devices	98
	References	99

5 Defects in Monocrystalline Silicon

<i>Wilfried von Ammon</i>	101
5.1 Technological Impact of Intrinsic Point Defects Aggregates	102
5.2 Thermophysical Properties of Intrinsic Point Defects	103
5.3 Aggregates of Intrinsic Point Defects	104
5.3.1 Experimental Observations	104
5.3.2 Theoretical Model: Incorporation of Intrinsic Point Defects	107
5.3.3 Theoretical Model: Aggregation of Intrinsic Point Defects	109
5.3.4 Effect of Impurities on Intrinsic Point Defect Aggregation	112
5.4 Formation of OSF Ring	115
References	117

6 Diffusion in Semiconductors

<i>Derek Shaw</i>	121
6.1 Basic Concepts	122
6.2 Diffusion Mechanisms	122
6.2.1 Vacancy and Interstitial Diffusion Mechanisms	122
6.2.2 The Interstitial–Substitutional Mechanism: Dissociative and Kick-Out Mechanisms	122
6.2.3 The Percolation Mechanism	123
6.3 Diffusion Regimes	123
6.3.1 Chemical Equilibrium: Self- and Isoconcentration Diffusion	123
6.3.2 Chemical Diffusion (or Diffusion Under Nonequilibrium Conditions)	123
6.3.3 Recombination–Enhanced Diffusion	125
6.3.4 Surface Effects	125
6.3.5 Short Circuit Paths	125
6.4 Internal Electric Fields	126
6.5 Measurement of Diffusion Coefficients	126
6.5.1 Anneal Conditions	126
6.5.2 Diffusion Sources	126
6.5.3 Profiling Techniques	126
6.5.4 Calculating the Diffusion Coefficient	127
6.6 Hydrogen in Semiconductors	127
6.7 Diffusion in Group IV Semiconductors	128
6.7.1 Germanium	128
6.7.2 Silicon	128
6.7.3 $\text{Si}_{1-x}\text{Ge}_x$ Alloys	129
6.7.4 Silicon Carbide	129
6.8 Diffusion in III–V Compounds	130
6.8.1 Self-Diffusion	130
6.8.2 Dopant Diffusion	130
6.8.3 Compositional Interdiffusion	131

6.9	Diffusion in II–VI Compounds	131
6.9.1	Self-Diffusion	132
6.9.2	Chemical Self-Diffusion	132
6.9.3	Dopant Diffusion	132
6.9.4	Compositional Interdiffusion	132
6.10	Conclusions	133
6.11	General Reading and References	133
	References	133
7	Photoconductivity in Materials Research	
	<i>Monica Brinza, Jan Willekens, Mohammed L. Benkhedir,</i>	
	<i>Guy J. Adriaenssens</i>	137
7.1	Steady State Photoconductivity Methods	138
7.1.1	The Basic Single-Beam Experiment	138
7.1.2	The Constant Photocurrent Method (CPM)	141
7.1.3	Dual-Beam Photoconductivity (DBP)	141
7.1.4	Modulated Photoconductivity (MPC)	141
7.2	Transient Photoconductivity Experiments	142
7.2.1	Current Relaxation from the Steady State	143
7.2.2	Transient Photoconductivity (TPC)	143
7.2.3	Time-of-Flight Measurements (TOF)	144
7.2.4	Interrupted Field Time-of-Flight (IFTOF)	145
	References	146
8	Electronic Properties of Semiconductor Interfaces	
	<i>Winfried Mönch</i>	147
8.1	Experimental Database	149
8.1.1	Barrier Heights of Laterally Homogeneous Schottky Contacts	149
8.1.2	Band Offsets of Semiconductor Heterostructures	152
8.2	IFIGS-and-Electronegativity Theory	153
8.3	Comparison of Experiment and Theory	155
8.3.1	Barrier Heights of Schottky Contacts	155
8.3.2	Band Offsets of Semiconductor Heterostructures	156
8.3.3	Band-Structure Lineup at Insulator Interfaces	158
8.4	Final Remarks	159
	References	159
9	Charge Transport in Disordered Materials	
	<i>Sergei Baranovskii, Oleg Rubel</i>	161
9.1	General Remarks on Charge Transport in Disordered Materials	163
9.2	Charge Transport in Disordered Materials via Extended States	167
9.3	Hopping Charge Transport in Disordered Materials	
	via Localized States	169
9.3.1	Nearest-Neighbor Hopping	170
9.3.2	Variable-Range Hopping	172

9.3.3 Description of Charge–Carrier Energy Relaxation and Hopping Conduction in Inorganic Noncrystalline Materials	173
9.3.4 Description of Charge Carrier Energy Relaxation and Hopping Conduction in Organic Noncrystalline Materials	180
9.4 Concluding Remarks	184
References	185

10 Dielectric Response

<i>Leonard Dissado</i>	187
10.1 Definition of Dielectric Response	188
10.1.1 Relationship to Capacitance	188
10.1.2 Frequency–Dependent Susceptibility	188
10.1.3 Relationship to Refractive Index	189
10.2 Frequency–Dependent Linear Responses	190
10.2.1 Resonance Response	190
10.2.2 Relaxation Response	192
10.3 Information Contained in the Relaxation Response	196
10.3.1 The Dielectric Increment for a Linear Response χ_0	196
10.3.2 The Characteristic Relaxation Time (Frequency)	199
10.3.3 The Relaxation Peak Shape	205
10.4 Charge Transport	208
10.5 A Few Final Comments	211
References	211

11 Ionic Conduction and Applications

<i>Harry L. Tuller</i>	213
11.1 Conduction in Ionic Solids	214
11.2 Fast Ion Conduction	216
11.2.1 Structurally Disordered Crystalline Solids	216
11.2.2 Amorphous Solids	219
11.2.3 Heavily Doped Defective Solids	219
11.2.4 Interfacial Ionic Conduction and Nanostructural Effects	220
11.3 Mixed Ionic–Electronic Conduction	221
11.3.1 Defect Equilibria	221
11.3.2 Electrolytic Domain Boundaries	222
11.4 Applications	223
11.4.1 Sensors	223
11.4.2 Solid Oxide Fuel Cells (SOFC)	224
11.4.3 Membranes	225
11.4.4 Batteries	225
11.4.5 Electrochromic Windows	226
11.5 Future Trends	226
References	226

Part B Growth and Characterization

12 Bulk Crystal Growth – Methods and Materials

<i>Peter Capper</i>	231
12.1 History	232
12.2 Techniques	233
12.2.1 Verneuil	233
12.2.2 Czochralski	233
12.2.3 Kyropoulos	234
12.2.4 Stepanov	235
12.2.5 Edge-Defined Film Growth	235
12.2.6 Bridgman	236
12.2.7 Vertical Gradient Freeze	237
12.2.8 Float Zone	237
12.2.9 Travelling Heater Method (THM)	237
12.2.10 Low-Temperature Solution Growth	238
12.2.11 High-Temperature Solution Growth (Flux)	238
12.2.12 Hydrothermal	239
12.2.13 Growth from the Vapor	240
12.3 Materials Grown	240
12.3.1 Group IV	240
12.3.2 Groups III–V	242
12.3.3 Groups II–VI	245
12.3.4 Oxides/Halides/Phosphates/Borates	249
12.4 Conclusions	251
References	251

13 Single-Crystal Silicon: Growth and Properties

<i>Fumio Shimura</i>	255
13.1 Overview	256
13.2 Starting Materials	257
13.2.1 Metallurgical-Grade Silicon	257
13.2.2 Polycrystalline Silicon	257
13.3 Single-Crystal Growth	258
13.3.1 Floating-Zone Method	259
13.3.2 Czochralski Method	261
13.3.3 Impurities in Czochralski Silicon	264
13.4 New Crystal Growth Methods	266
13.4.1 Czochralski Growth with an Applied Magnetic Field (MCZ) ...	266
13.4.2 Continuous Czochralski Method (CCZ)	267
13.4.3 Neckingless Growth Method	267
References	268

14 Epitaxial Crystal Growth: Methods and Materials

<i>Peter Capper, Stuart Irvine, Tim Joyce</i>	271
14.1 Liquid-Phase Epitaxy (LPE)	271
14.1.1 Introduction and Background	271

14.1.2	History and Status	272
14.1.3	Characteristics	272
14.1.4	Apparatus and Techniques	273
14.1.5	Group IV	275
14.1.6	Group III–V	276
14.1.7	Group II–VI	278
14.1.8	Atomically Flat Surfaces	280
14.1.9	Conclusions	280
14.2	Metalorganic Chemical Vapor Deposition (MOCVD)	280
14.2.1	Introduction and Background	280
14.2.2	Basic Reaction Kinetics	281
14.2.3	Precursors	283
14.2.4	Reactor Cells	284
14.2.5	III–V MOCVD	286
14.2.6	II–VI MOCVD	288
14.2.7	Conclusions	290
14.3	Molecular Beam Epitaxy (MBE)	290
14.3.1	Introduction and Background	290
14.3.2	Reaction Mechanisms	291
14.3.3	MBE Growth Systems	293
14.3.4	Gas Sources in MBE	295
14.3.5	Growth of III–V Materials by MBE	296
14.3.6	Conclusions	299
	References	299

15 Narrow-Bandgap II–VI Semiconductors: Growth

	<i>Peter Capper</i>	303
15.1	Bulk Growth Techniques	304
15.1.1	Phase Equilibria	304
15.1.2	Crystal Growth	304
15.1.3	Material Characterization	306
15.2	Liquid-Phase Epitaxy (LPE)	308
15.2.1	Hg-Rich Growth	309
15.2.2	Te-Rich Growth	309
15.2.3	Material Characteristics	311
15.3	Metalorganic Vapor Phase Epitaxy (MOVPE)	312
15.3.1	Substrate Type and Orientation	315
15.3.2	Doping	316
15.3.3	In Situ Monitoring	317
15.4	Molecular Beam Epitaxy (MBE)	317
15.4.1	Double-Layer Heterojunction Structures	319
15.4.2	Multilayer Heterojunction Structures	319
15.4.3	CMT and CdZnTe Growth on Silicon	319
15.5	Alternatives to CMT	320
	References	321

16 Wide-Bandgap II–VI Semiconductors: Growth and Properties

<i>Jifeng Wang, Minoru Isshiki</i>	325
16.1 Crystal Properties	326
16.1.1 Basic Properties	326
16.1.2 Phase Diagram	326
16.2 Epitaxial Growth	328
16.2.1 The LPE Technique	329
16.2.2 Vapor–Phase Epitaxy Techniques	330
16.3 Bulk Crystal Growth	333
16.3.1 The CVT and PVT Techniques	333
16.3.2 Hydrothermal Growth	336
16.3.3 Bridgman and Gradient Freezing (GF) Method	337
16.3.4 The Traveling Heater Method (THM)	338
16.3.5 Other Methods	338
16.4 Conclusions	339
References	340

17 Structural Characterization

<i>Paul D. Brown</i>	343
17.1 Radiation–Material Interactions	344
17.2 Particle–Material Interactions	345
17.3 X-Ray Diffraction	348
17.4 Optics, Imaging and Electron Diffraction	351
17.4.1 Electron Diffraction and Image Contrast Analysis	355
17.4.2 Microdiffraction and Polarity	358
17.4.3 Reflection High–Energy Electron Diffraction	359
17.5 Characterizing Functional Activity	362
17.6 Sample Preparation	362
17.7 Case Studies – Complementary Characterization of Electronic and Optoelectronic Materials	364
17.7.1 Identifying Defect Sources Within Homoepitaxial GaN	366
17.7.2 Cathodoluminescence/Correlated TEM Investigation of Epitaxial GaN	367
17.7.3 Scanning Transmission Electron Beam Induced Conductivity of Si/Si _{1–x} Ge _x /Si(001)	367
17.8 Concluding Remarks	370
References	370

18 Surface Chemical Analysis

<i>David Sykes</i>	373
18.1 Electron Spectroscopy	373
18.1.1 Auger Electron Spectroscopy	373
18.1.2 X-Ray Photoelectron Spectroscopy (XPS)	375
18.2 Glow–Discharge Spectroscopies (GDOES and GDMS)	376
18.3 Secondary Ion Mass Spectrometry (SIMS)	377
18.4 Conclusion	384

19 Thermal Properties and Thermal Analysis: Fundamentals, Experimental Techniques and Applications	
<i>Safa Kasap, Dan Tonchev</i>	385
19.1 Heat Capacity	386
19.1.1 Fundamental Debye Heat Capacity of Crystals	386
19.1.2 Specific Heat Capacity of Selected Groups of Materials	388
19.2 Thermal Conductivity	391
19.2.1 Definition and Typical Values	391
19.2.2 Thermal Conductivity of Crystalline Insulators	391
19.2.3 Thermal Conductivity of Noncrystalline Insulators	393
19.2.4 Thermal Conductivity of Metals	395
19.3 Thermal Expansion	396
19.3.1 Grüneisen's Law and Anharmonicity	396
19.3.2 Thermal Expansion Coefficient α	398
19.4 Enthalpic Thermal Properties	398
19.4.1 Enthalpy, Heat Capacity and Physical Transformations	398
19.4.2 Conventional Differential Scanning Calorimetry (DSC)	400
19.5 Temperature-Modulated DSC (TMDSC)	403
19.5.1 TMDSC Principles	403
19.5.2 TMDSC Applications	404
19.5.3 Tzero Technology	405
References	406
 20 Electrical Characterization of Semiconductor Materials and Devices	
<i>M. Jamal Deen, Fabien Pascal</i>	409
20.1 Resistivity	410
20.1.1 Bulk Resistivity	410
20.1.2 Contact Resistivity	415
20.2 Hall Effect	418
20.2.1 Physical Principles	419
20.2.2 Hall Scattering Factor	420
20.3 Capacitance-Voltage Measurements	421
20.3.1 Average Doping Density by Maximum-Minimum High-Frequency Capacitance Method	421
20.3.2 Doping Profile by High-Frequency and High-Low Frequency Capacitance Methods	422
20.3.3 Density of Interface States	424
20.4 Current-Voltage Measurements	426
20.4.1 I - V Measurements on a Simple Diode	426
20.4.2 I - V Measurements on a Simple MOSFET	426
20.4.3 Floating Gate Measurements	427
20.5 Charge Pumping	428
20.6 Low-Frequency Noise	430
20.6.1 Introduction	430
20.6.2 Noise from the Interfacial Oxide Layer	431

20.6.3 Impedance Considerations During Noise Measurement	432
20.7 Deep-Level Transient Spectroscopy	434
References	436

Part C Materials for Electronics

21 Single-Crystal Silicon: Electrical and Optical Properties

<i>Shlomo Hava, Mark Auslender</i>	441
21.1 Silicon Basics	441
21.1.1 Structure and Energy Bands	441
21.1.2 Impurity Levels and Charge-Carrier Population	443
21.1.3 Carrier Concentration, Electrical and Optical Properties	446
21.1.4 Theory of Electrical and Optical Properties	447
21.2 Electrical Properties	451
21.2.1 Ohm's Law Regime	451
21.2.2 High-Electric-Field Effects	465
21.2.3 Review Material	471
21.3 Optical Properties	472
21.3.1 Diversity of Silicon as an Optical Material	472
21.3.2 Measurements of Optical Constants	472
21.3.3 Modeling of Optical Constants	474
21.3.4 Electric-Field and Temperature Effects on Optical Constants	477
References	478

22 Silicon-Germanium: Properties, Growth and Applications

<i>Peter Ashburn, Darren M. Bagnall</i>	481
22.1 Physical Properties of Silicon-Germanium	482
22.1.1 Critical Thickness	482
22.1.2 Band Structure	483
22.1.3 Dielectric Constant	484
22.1.4 Density of States	484
22.1.5 Majority-Carrier Mobility in Strained $\text{Si}_{1-x}\text{Ge}_x$	486
22.1.6 Majority-Carrier Mobility in Tensile-Strained Si on Relaxed $\text{Si}_{1-x}\text{Ge}_x$	486
22.1.7 Minority-Carrier Mobility in Strained $\text{Si}_{1-x}\text{Ge}_x$	486
22.1.8 Apparent Band-Gap Narrowing in $\text{Si}_{1-x}\text{Ge}_x$ HBTs	487
22.2 Optical Properties of SiGe	488
22.2.1 Dielectric Functions and Interband Transitions	488
22.2.2 Photoluminescence	489
22.2.3 SiGe Quantum Wells	490
22.3 Growth of Silicon-Germanium	492
22.3.1 In-Situ Hydrogen Bake	492
22.3.2 Hydrogen Passivation	492
22.3.3 Ultra-Clean Epitaxy Systems	492

22.3.4	$\text{Si}_{1-x}\text{Ge}_x$ Epitaxy	492
22.3.5	Selective $\text{Si}_{1-x}\text{Ge}_x$ Epitaxy	492
22.4	Polycrystalline Silicon–Germanium	494
22.4.1	Electrical Properties of Polycrystalline $\text{Si}_{1-x}\text{Ge}_x$	496
References	497

23 Gallium Arsenide

<i>Mike Brozel</i>	499
23.1 Bulk Growth of GaAs	502
23.1.1 Doping Considerations	502
23.1.2 Horizontal Bridgman and Horizontal Gradient Freeze Techniques	503
23.1.3 Liquid–Encapsulated Czochralski (LEC) Technique	504
23.1.4 Vertical Gradient Freeze (VGF) Technique	506
23.2 Epitaxial Growth of GaAs	507
23.2.1 Liquid–Phase Epitaxy (LPE)	507
23.2.2 Vapour–Phase Epitaxy (VPE) Technologies	508
23.2.3 Molecular–Beam Epitaxy (MBE)	509
23.2.4 Growth of Epitaxial and Pseudomorphic Structures	511
23.3 Diffusion in Gallium Arsenide	511
23.3.1 Shallow Acceptors	512
23.3.2 Shallow Donors	513
23.3.3 Transition Metals	513
23.4 Ion Implantation into GaAs	513
23.5 Crystalline Defects in GaAs	514
23.5.1 Defects in Melt–Grown GaAs	514
23.5.2 Epitaxial GaAs (not Low Temperature MBE GaAs)	516
23.5.3 LTMBE GaAs	517
23.6 Impurity and Defect Analysis of GaAs (Chemical)	517
23.7 Impurity and Defect Analysis of GaAs (Electrical)	518
23.7.1 Introduction to Electrical Analysis of Defects in GaAs	518
23.8 Impurity and Defect Analysis of GaAs (Optical)	521
23.8.1 Optical Analysis of Defects in GaAs	521
23.9 Assessment of Complex Heterostructures	522
23.9.1 Carrier Concentration Measurements in Heterostructures	522
23.9.2 Layer Thickness and Composition Measurements	522
23.10 Electrical Contacts to GaAs	524
23.10.1 Ohmic Contacts	524
23.10.2 Schottky Contacts	524
23.11 Devices Based on GaAs (Microwave)	524
23.11.1 The Gunn Diode	524
23.11.2 The Metal–Semiconductor Field–Effect Transistor (MESFET)	525
23.11.3 The High–Electron–Mobility Transistor (HEMT) or Modulation Doped FET (MODFET)	526
23.11.4 The Heterojunction Bipolar Transistor (HBT)	526

23.12	Devices based on GaAs (Electro-optical)	527
23.12.1	GaAs Emitters	527
23.12.2	GaAs Modulators	531
23.12.3	GaAs Photodetectors	531
23.13	Other Uses for GaAs	532
23.14	Conclusions	532
	References	533
24	High-Temperature Electronic Materials: Silicon Carbide and Diamond	
	<i>Magnus Willander, Milan Friesel, Qamar-ul Wahab, Boris Straumal</i>	537
24.1	Material Properties and Preparation	540
24.1.1	Silicon Carbide	540
24.1.2	Diamond	544
24.2	Electronic Devices	547
24.2.1	Silicon Carbide	547
24.2.2	Diamond	551
24.3	Summary	557
	References	558
25	Amorphous Semiconductors: Structure, Optical, and Electrical Properties	
	<i>Kazuo Morigaki, Chisato Ogihara</i>	565
25.1	Electronic States	565
25.2	Structural Properties	568
25.2.1	General Aspects	568
25.2.2	a-Si:H and Related Materials	568
25.2.3	Chalcogenide Glasses	569
25.3	Optical Properties	570
25.3.1	General Aspects	570
25.3.2	a-Si:H and Related Materials	571
25.3.3	Chalcogenide Glasses	572
25.4	Electrical Properties	573
25.4.1	General Aspects	573
25.4.2	a-Si:H and Related Materials	574
25.4.3	Chalcogenide Glasses	575
25.5	Light-Induced Phenomena	575
25.6	Nanosized Amorphous Structure	577
	References	578
26	Amorphous and Microcrystalline Silicon	
	<i>Akihisa Matsuda</i>	581
26.1	Reactions in SiH ₄ and SiH ₄ /H ₂ Plasmas	581
26.2	Film Growth on a Surface	583
26.2.1	Growth of a-Si:H	583
26.2.2	Growth of μ c-Si:H	584

26.3	Defect Density Determination for a-Si:H and $\mu\text{c-Si:H}$	589
26.3.1	Dangling Bond Defects	589
26.3.2	Dangling Bond Defect Density in $\mu\text{c-Si:H}$	590
26.4	Device Applications	590
26.5	Recent Progress in Material Issues Related to Thin-Film Silicon Solar Cells	591
26.5.1	Controlling Photoinduced Degradation in a-Si:H	591
26.5.2	High Growth Rates of Device-Grade $\mu\text{c-Si:H}$	592
26.6	Summary	594
	References	594

27 Ferroelectric Materials

	<i>Roger Whatmore</i>	597
27.0.1	Definitions and Background	597
27.0.2	Basic Ferroelectric Characteristics and Models	599
27.1	Ferroelectric Materials	601
27.1.1	Ferroelectric Oxides	601
27.1.2	Triglycine Sulphate (TGS)	607
27.1.3	Polymeric Ferroelectrics	607
27.2	Ferroelectric Materials Fabrication Technology	608
27.2.1	Single Crystals	608
27.2.2	Ceramics	609
27.2.3	Thick Films	613
27.2.4	Thin Films	613
27.3	Ferroelectric Applications	616
27.3.1	Dielectrics	616
27.3.2	Computer Memories	616
27.3.3	Piezoelectrics	617
27.3.4	Pyroelectrics	620
	References	622

28 Dielectric Materials for Microelectronics

	<i>Robert M. Wallace</i>	625
28.0.1	The Scaling of Integrated Circuits	625
28.0.2	Role of Dielectrics for ICs	629
28.1	Gate Dielectrics	630
28.1.1	Transistor Structure	630
28.1.2	Transistor Dielectric Requirements in View of Scaling	630
28.1.3	Silicon Dioxide	635
28.1.4	Silicon Oxynitride: SiO_xN_y	641
28.1.5	High- κ Dielectrics	643
28.2	Isolation Dielectrics	647
28.3	Capacitor Dielectrics	647
28.3.1	Types of IC Memory	647
28.3.2	Capacitor Dielectric Requirements in View of Scaling	648
28.3.3	Dielectrics for Volatile Memory Capacitors	648
28.3.4	Dielectrics for Nonvolatile Memory	649

28.4 Interconnect Dielectrics 651

28.4.1 Tetraethoxysilane (TEOS) 651

28.4.2 Low- κ Dielectrics 651

28.5 Summary 653

References 653

29 Thin Films

Robert D. Gould† 659

29.1 Deposition Methods 661

29.1.1 Physical Deposition Methods 661

29.1.2 Chemical Deposition Methods 677

29.2 Structure 682

29.2.1 Crystallography 682

29.2.2 Film Structure 683

29.2.3 Morphology 688

29.3 Properties 692

29.3.1 Optical Properties 692

29.3.2 Electrical Properties 696

29.4 Concluding Remarks 708

References 711

30 Thick Films

Neil White 717

30.1 Thick Film Processing 718

30.1.1 Screen Printing 718

30.1.2 The Drying and Firing Process 719

30.2 Substrates 720

30.2.1 Alumina 720

30.2.2 Stainless Steel 720

30.2.3 Polymer Substrates 720

30.3 Thick Film Materials 721

30.3.1 Conductors 721

30.3.2 Resistors 722

30.3.3 Dielectrics 723

30.3.4 Polymer Thick Films 723

30.4 Components and Assembly 724

30.4.1 Passive Components 724

30.4.2 Active Components 725

30.4.3 Trimming 725

30.4.4 Wire Bonding 726

30.4.5 Soldering of Surface-Mounted Components 727

30.4.6 Packaging and Testing 727

30.5 Sensors 728

30.5.1 Mechanical 728

30.5.2 Thermal 729

30.5.3 Optical 730

30.5.4 Chemical	730
30.5.5 Magnetic	730
30.5.6 Actuators	731
References	731

Part D Materials for Optoelectronics and Photonics

31 III–V Ternary and Quaternary Compounds

<i>Sadao Adachi</i>	735
31.1 Introduction to III–V Ternary and Quaternary Compounds	735
31.2 Interpolation Scheme	736
31.3 Structural Parameters	737
31.3.1 Lattice Parameters and Lattice-Matching Conditions Between III–V Quaternaries and Binary Substrates	737
31.3.2 Molecular and Crystal Densities	737
31.4 Mechanical, Elastic and Lattice Vibronic Properties	739
31.4.1 Microhardness	739
31.4.2 Elastic Constants and Related Moduli	739
31.4.3 Long-Wavelength Phonons	739
31.5 Thermal Properties	741
31.5.1 Specific Heat and Debye Temperature	741
31.5.2 Thermal Expansion Coefficient	741
31.5.3 Thermal Conductivity	741
31.6 Energy Band Parameters	743
31.6.1 Bandgap Energy	743
31.6.2 Carrier Effective Mass	744
31.6.3 Deformation Potential	746
31.7 Optical Properties	748
31.7.1 The Reststrahlen Region	748
31.7.2 The Interband Transition Region	749
31.8 Carrier Transport Properties	750
References	751

32 Group III Nitrides

<i>Ali Teke, Hadis Morkoç</i>	753
32.1 Crystal Structures of Nitrides	755
32.2 Lattice Parameters of Nitrides	756
32.3 Mechanical Properties of Nitrides	757
32.4 Thermal Properties of Nitrides	761
32.4.1 Thermal Expansion Coefficients	761
32.4.2 Thermal Conductivity	762
32.4.3 Specific Heat	764
32.5 Electrical Properties of Nitrides	766
32.5.1 Low-Field Transport	766
32.5.2 High-Field Transport	775

32.6	Optical Properties of Nitrides	777
32.6.1	Gallium Nitride	778
32.6.2	Aluminium Nitride	786
32.6.3	Indium Nitride	789
32.7	Properties of Nitride Alloys	791
32.8	Summary and Conclusions	794
	References	795

33 **Electron Transport Within the III–V Nitride Semiconductors, GaN, AlN, and InN: A Monte Carlo Analysis**

Brian E. Foutz, Stephen K. O’Leary, Michael Shur, Lester F. Eastman 805

33.1	Electron Transport Within Semiconductors and the Monte Carlo Simulation Approach	806
33.1.1	The Boltzmann Transport Equation	807
33.1.2	Our Ensemble Semi-Classical Monte Carlo Simulation Approach	808
33.1.3	Parameter Selections for Bulk Wurtzite GaN, AlN, and InN	808
33.2	Steady-State and Transient Electron Transport Within Bulk Wurtzite GaN, AlN, and InN	810
33.2.1	Steady-State Electron Transport Within Bulk Wurtzite GaN ..	811
33.2.2	Steady-State Electron Transport: A Comparison of the III–V Nitride Semiconductors with GaAs	812
33.2.3	Influence of Temperature on the Electron Drift Velocities Within GaN and GaAs	812
33.2.4	Influence of Doping on the Electron Drift Velocities Within GaN and GaAs	815
33.2.5	Electron Transport in AlN	816
33.2.6	Electron Transport in InN	818
33.2.7	Transient Electron Transport	820
33.2.8	Electron Transport: Conclusions	822
33.3	Electron Transport Within III–V Nitride Semiconductors: A Review ...	822
33.3.1	Evolution of the Field	822
33.3.2	Recent Developments	824
33.3.3	Future Perspectives	825
33.4	Conclusions	826
	References	826

34 **II–IV Semiconductors for Optoelectronics: CdS, CdSe, CdTe**

Jifeng Wang, Minoru Isshiki 829

34.1	Background	829
34.2	Solar Cells	829
34.2.1	Basic Description of Solar Cells	829
34.2.2	Design of Cd–Based Solar Cells	830
34.2.3	Development of CdS/CdTe Solar Cells	831
34.2.4	CdZnTe Solar Cells	834
34.2.5	The Future of Cd–Based Solar Cells	834

34.3	Radiation Detectors	834
34.3.1	Basic Description of Semiconductor Radiation Detectors	835
34.3.2	CdTe and CdZnTe Radiation Detectors	835
34.3.3	Performance of CdTe and CdZnTe Detectors	836
34.3.4	Applications of CdTe and CdZnTe Detectors.....	839
34.4	Conclusions	840
	References	840
35	Doping Aspects of Zn-Based Wide-Band-Gap Semiconductors	
	<i>Gertrude F. Neumark, Yinyan Gong, Igor L. Kuskovsky</i>	843
35.1	ZnSe	843
35.1.1	Doping – Overview	843
35.1.2	Results on p-Type Material with N as the Primary Dopant ..	845
35.2	ZnBeSe.....	848
35.3	ZnO	849
35.3.1	Doping	849
35.3.2	Optical Properties	850
	References	851
36	II–VI Narrow-Bandgap Semiconductors for Optoelectronics	
	<i>Ian M. Baker</i>	855
36.0.1	Historical Perspective and Early Detectors	856
36.0.2	Introduction to HgCdTe	857
36.0.3	Introduction to Device Types	857
36.1	Applications and Sensor Design	858
36.2	Photoconductive Detectors in HgCdTe and Related Alloys.....	860
36.2.1	Introduction to the Technology of Photoconductor Arrays ...	860
36.2.2	Theoretical Fundamentals for Long-Wavelength Arrays	861
36.2.3	Special Case of Medium-Wavelength Arrays	863
36.2.4	Nonequilibrium Effects in Photoconductors	863
36.3	SPRITE Detectors	864
36.4	Photoconductive Detectors in Closely Related Alloys	866
36.5	Conclusions on Photoconductive HgCdTe Detectors	867
36.6	Photovoltaic Devices in HgCdTe	867
36.6.1	Ideal Photovoltaic Devices	868
36.6.2	Nonideal Behavior in HgCdTe Diodes	869
36.6.3	Theoretical Foundations of HgCdTe Array Technology	870
36.6.4	Manufacturing Technology for HgCdTe Arrays.....	873
36.6.5	HgCdTe 2-D Arrays for the 3–5 μm (MW) Band	878
36.6.6	HgCdTe 2-D Arrays for the 8–12 μm (LW) Band	879
36.6.7	HgCdTe 2-D Arrays for the 1–3 μm (SW) Band	879
36.6.8	Towards “GEN III Detectors”	880
36.6.9	Conclusions and Future Trends for Photovoltaic HgCdTe Arrays	882
36.7	Emission Devices in II–VI Semiconductors	882
36.8	Potential for Reduced-Dimensionality HgTe–CdTe	883
	References	883

37 Optoelectronic Devices and Materials	
<i>Stephen Sweeney, Alfred Adams</i>	887
37.1 Introduction to Optoelectronic Devices	888
37.1.1 Historical Perspective	888
37.2 Light-Emitting Diodes and Semiconductor Lasers	890
37.2.1 Carrier-Photon Interactions in Semiconductors	890
37.2.2 Direct- and Indirect-Gap Semiconductors	890
37.2.3 Emission and Absorption Rates and the Einstein Relations	891
37.2.4 Population Inversion	892
37.2.5 Gain in Semiconductors	892
37.2.6 Density of States	893
37.2.7 Optical Feedback in a Fabry-Perot Laser Cavity	896
37.2.8 Wave-Guiding	897
37.2.9 Carrier Confinement	898
37.2.10 Current Confinement	898
37.2.11 Laser Threshold and Efficiency	899
37.2.12 Carrier Recombination Processes	900
37.2.13 Temperature Sensitivity and T_0	903
37.3 Single-Mode Lasers	904
37.3.1 DFB lasers	904
37.3.2 VCSELs	905
37.4 Optical Amplifiers	906
37.4.1 An Introduction to Optical Amplification	906
37.4.2 Semiconductor Optical Amplifiers (SOAs)	907
37.5 Modulators	907
37.5.1 Modulator Theory	907
37.5.2 Polarisation-Insensitive Modulators	909
37.5.3 High-Speed High-Power QCSE Modulators	910
37.5.4 The Electro-Optic Effect	911
37.6 Photodetectors	911
37.6.1 Photodetector Requirements	912
37.6.2 Photodetection Theory	912
37.6.3 Detectors with Internal Gain	913
37.6.4 Avalanche Photodetectors	913
37.7 Conclusions	914
References	915
38 Liquid Crystals	
<i>David Dunmur, Geoffrey Luckhurst</i>	917
38.1 Introduction to Liquid Crystals	917
38.1.1 Calamitic Liquid Crystals	919
38.1.2 Chiral Liquid Crystals	921
38.1.3 Discotic Liquid Crystals	923
38.2 The Basic Physics of Liquid Crystals	924
38.2.1 Orientational Order	924
38.2.2 Director Alignment	925

38.2.3	Elasticity	926
38.2.4	Flexoelectricity	928
38.2.5	Viscosity	929
38.3	Liquid-Crystal Devices	931
38.3.1	A Model Liquid-Crystal Display: Electrically Controlled Birefringence Mode (ECB)	932
38.3.2	High-Volume Commercial Displays: The Twisted Nematic (TN) and Super-Twisted Nematic (STN) Displays	935
38.3.3	Complex LC Displays and Other Cell Configurations	937
38.4	Materials for Displays	940
38.4.1	Chemical Structure and Liquid-Crystal Phase Behaviour	942
38.4.2	The Formulation of Liquid-Crystal Display Mixtures	942
38.4.3	Relationships Between Physical Properties and Chemical Structures of Mesogens	943
	References	949
39	Organic Photoconductors	
	<i>David S. Weiss, Martin Abkowitz</i>	953
39.1	Chester Carlson and Xerography	954
39.2	Operational Considerations and Critical Materials Properties	956
39.2.1	Dark Conductivity	956
39.2.2	Photodischarge-Charge Transport	957
39.2.3	Photogeneration	963
39.3	OPC Characterization	965
39.3.1	Dark Decay	965
39.3.2	Photosensitivity	965
39.3.3	Electrical-Only Cycling	966
39.4	OPC Architecture and Composition	967
39.4.1	OPC Architecture	967
39.4.2	Coating Technologies	968
39.4.3	Substrate and Conductive Layer	969
39.4.4	Smoothing Layer and Charge-Blocking Layer	969
39.4.5	Charge-Generation Layer (CGL)	970
39.4.6	Charge-Transport Layer (CTL)	974
39.4.7	Backing Layer	975
39.4.8	Overcoat Layer	975
39.5	Photoreceptor Fabrication	976
39.6	Summary	977
	References	978
40	Luminescent Materials	
	<i>Andy Edgar</i>	983
40.1	Luminescent Centres	985
40.1.1	Rare-Earth Ions	985
40.1.2	Transition-Metal Ions	986
40.1.3	s^2 Ions	987
40.1.4	Semiconductors	987

40.2	Interaction with the Lattice	987
40.3	Thermally Stimulated Luminescence	989
40.4	Optically (Photo-)Stimulated Luminescence	990
40.5	Experimental Techniques – Photoluminescence	991
40.6	Applications	992
40.6.1	White Light-Emitting Diodes (LEDs)	992
40.6.2	Long-Persistence Phosphors	992
40.6.3	X-Ray Storage Phosphors	993
40.6.4	Phosphors for Optical Displays	994
40.6.5	Scintillators	994
40.7	Representative Phosphors	995
	References	995

41 Nano-Engineered Tunable Photonic Crystals in the Near-IR and Visible Electromagnetic Spectrum

	<i>Harry Ruda, Naomi Matsuura</i>	997
41.1	PC Overview	998
41.1.1	Introduction to PCs	998
41.1.2	Nano-Engineering of PC Architectures	999
41.1.3	Materials Selection for PCs	1000
41.2	Traditional Fabrication Methodologies for Static PCs	1001
41.2.1	2-D PC Structures	1001
41.2.2	3-D PC Structures	1007
41.3	Tunable PCs	1011
41.3.1	Tuning the PC Response by Changing the Refractive Index of the Constituent Materials	1011
41.3.2	Tuning PC Response by Altering the Physical Structure of the PC	1012
41.4	Summary and Conclusions	1014
	References	1015

42 Quantum Wells, Superlattices, and Band-Gap Engineering

	<i>Mark Fox</i>	1021
42.1	Principles of Band-Gap Engineering and Quantum Confinement	1022
42.1.1	Lattice Matching	1022
42.1.2	Quantum-Confined Structures	1023
42.2	Optoelectronic Properties of Quantum-Confined Structures	1024
42.2.1	Electronic States in Quantum Wells and Superlattices	1024
42.2.2	Interband Optical Transitions	1026
42.2.3	The Quantum-Confined Stark Effect	1028
42.2.4	Inter-Sub-Band Transitions	1029
42.2.5	Vertical Transport	1030
42.2.6	Carrier Capture and Relaxation	1031
42.3	Emitters	1032
42.3.1	Interband Light-Emitting Diodes and Lasers	1032
42.3.2	Quantum Cascade Lasers	1033

42.4	Detectors	1034
42.4.1	Solar Cells	1034
42.4.2	Avalanche Photodiodes	1034
42.4.3	Inter-Sub-Band Detectors	1035
42.4.4	Unipolar Avalanche Photodiodes	1035
42.5	Modulators	1036
42.6	Future Directions	1037
42.7	Conclusions	1038
	References	1038

43 Glasses for Photonic Integration

	<i>Ray DeCorby</i>	1041
43.1	Main Attributes of Glasses as Photonic Materials	1042
43.1.1	The Glass Transition as Enabler	1043
43.1.2	Metastability	1046
43.1.3	Glass as Host Material	1049
43.2	Glasses for Integrated Optics	1050
43.2.1	Low Index Glassy Films	1050
43.2.2	Medium Index Glassy Films	1051
43.2.3	High Index Glassy Films	1051
43.3	Laser Glasses for Integrated Light Sources	1053
43.3.1	Advantages of Glass-based Light Sources	1053
43.3.2	Alternative Glass Hosts	1054
43.3.3	Progress Towards Integrated Light Sources in Glass	1056
43.4	Summary	1057
	References	1059

44 Optical Nonlinearity in Photonic Glasses

	<i>Keiji Tanaka</i>	1063
44.1	Third-Order Nonlinearity in Homogeneous Glass	1064
44.1.1	Experimental	1064
44.1.2	Theoretical Treatment	1065
44.1.3	Stimulated Light Scattering and Supercontinuum Generation	1068
44.2	Second-Order Nonlinearity in Poled Glass	1069
44.3	Particle-Embedded Systems	1070
44.4	Photoinduced Phenomena	1071
44.5	Summary	1072
	References	1072

45 Nonlinear Optoelectronic Materials

	<i>Lukasz Brzozowski, Edward Sargent</i>	1075
45.1	Background	1075
45.1.1	Signal Processing in Optical Networks	1075
45.1.2	Optical Signal Processing Using Nonlinear Optics	1076
45.1.3	The Approach Taken During this Survey of Nonlinear Optoelectronic Materials	1076

45.2	Illumination-Dependent Refractive Index and Nonlinear Figures of Merit (FOM)	1077
45.2.1	Ultrafast Response	1077
45.2.2	Ultrafast Nonlinear Material Figures of Merit	1078
45.2.3	Resonant Response	1079
45.2.4	Resonant Nonlinear Material Figures of Merit	1079
45.3	Bulk and Multi-Quantum-Well (MQW) Inorganic Crystalline Semiconductors	1080
45.3.1	Resonant Nonlinearities	1080
45.3.2	Nonresonant Nonlinearities in Inorganic Crystalline Semiconductors	1083
45.4	Organic Materials	1084
45.4.1	Resonant Nonlinear Response of Organic Materials	1085
45.4.2	Nonresonant Nonlinear Response of Organic Materials	1086
45.5	Nanocrystals	1087
45.6	Other Nonlinear Materials	1088
45.7	Conclusions	1089
	References	1089

Part E Novel Materials and Selected Applications

46	Solar Cells and Photovoltaics	
	<i>Stuart Irvine</i>	1095
46.1	Figures of Merit for Solar Cells	1096
46.2	Crystalline Silicon	1098
46.3	Amorphous Silicon	1100
46.4	GaAs Solar Cells	1101
46.5	CdTe Thin-Film Solar Cells	1102
46.6	CuInGaSe ₂ (CIGS) Thin-Film Solar Cells	1103
46.7	Conclusions	1104
	References	1105
47	Silicon on Mechanically Flexible Substrates for Large-Area Electronics	
	<i>Peyman Servati, Arokia Nathan</i>	1107
47.1	a-Si:H TFTs on Flexible Substrates	1108
47.2	Field-Effect Transport in Amorphous Films	1108
47.2.1	Localized and Extended States	1109
47.2.2	Density of States (DOS)	1110
47.2.3	Effective Carrier Mobility	1110
47.3	Electronic Transport Under Mechanical Stress	1113
47.3.1	Thin-Film Strain Gauges	1114
47.3.2	Strained Amorphous-Silicon Transistors	1116
	References	1118

48 Photoconductors for X-Ray Image Detectors	
<i>M. Zahangir Kabir, Safa Kasap, John Rowlands</i>	1121
48.1 X-Ray Photoconductors	1123
48.1.1 Ideal Photoconductor Properties	1123
48.1.2 Potential Photoconductors	1124
48.1.3 Summary and the Future	1130
48.2 Metrics of Detector Performance	1131
48.2.1 X-Ray Sensitivity	1131
48.2.2 Detective Quantum Efficiency	1133
48.2.3 Modulation Transfer Function (MTF)	1134
48.3 Conclusion	1136
References	1136
49 Phase-Change Optical Recording	
<i>Alexander Kolobov, Junji Tominaga</i>	1139
49.1 Digital Versatile Discs (DVDs)	1140
49.1.1 Questions Pertaining to DVD Recording	1140
49.1.2 Local Structure and its Changes During the Phase Transition	1140
49.1.3 Related Issues	1143
49.2 Super-RENS Discs	1144
49.3 In Lieu of Conclusion	1145
References	1145
50 Carbon Nanotubes and Bucky Materials	
<i>Mark Baxendale</i>	1147
50.1 Carbon Nanotubes	1147
50.1.1 General	1147
50.1.2 Geometry	1149
50.1.3 Synthesis and Chemistry	1149
50.1.4 Electronic Structure and Transport	1150
50.1.5 Nanoelectronic Devices	1151
50.1.6 Other Electronic Applications	1152
50.2 Bucky Materials	1153
References	1153
51 Magnetic Information-Storage Materials	
<i>Larry Comstock</i>	1155
51.1 Magnetic Recording Technology	1156
51.1.1 Magnetic Thin Films	1157
51.1.2 The Write Head	1158
51.1.3 Spin Valve Read Head	1162
51.1.4 Longitudinal Recording Media	1169
51.1.5 Perpendicular Recording	1175
51.2 Magnetic Random-Access Memory	1185
51.2.1 Tunneling Magnetoresistive Heads	1188

51.3	Extraordinary Magnetoresistance (EMR)	1189
51.4	Summary	1189
	References	1189

52 High-Temperature Superconductors

	<i>Rainer Wesche</i>	1193
52.1	The Superconducting State	1195
52.1.1	Characteristic Properties of Superconductors	1195
52.1.2	Superconductor Electrodynamics	1197
52.1.3	Superconductivity: A Macroscopic Quantum Phenomenon ..	1198
52.1.4	Type II Superconductors	1200
52.2	Cuprate High- T_c Superconductors: An Overview	1202
52.2.1	Major Families of Cuprate Superconductors	1202
52.2.2	Generic Phase Diagram of Cuprate Superconductors	1202
52.2.3	Crystal Structures	1204
52.2.4	Critical Temperatures	1205
52.3	Physical Properties of Cuprate Superconductors	1207
52.3.1	Anisotropic Superconductors	1207
52.3.2	Irreversibility Line	1208
52.3.3	Limitations of the Transport Critical Current	1209
52.4	Superconducting Films	1212
52.5	The Special Case of MgB_2	1214
52.6	Summary	1216
	References	1216

53 Molecular Electronics

	<i>Michael Petty</i>	1219
53.1	Electrically Conductive Organic Compounds	1220
53.1.1	Orbitals and Chemical Bonding	1220
53.1.2	Band Theory	1221
53.1.3	Electrical Conductivity	1222
53.2	Materials	1223
53.3	Plastic Electronics	1225
53.3.1	Diodes and Transistors	1225
53.3.2	Organic Light-Emitting Structures	1226
53.3.3	Photovoltaic Devices	1227
53.3.4	Chemical Sensors	1228
53.4	Molecular-Scale Electronics	1229
53.4.1	Moore's Laws	1229
53.4.2	Nanoscale Organic Films	1230
53.4.3	Patterning Technologies	1232
53.4.4	Molecular Device Architectures	1233
53.5	DNA Electronics	1235
53.6	Conclusions	1236
	References	1237

54 Organic Materials for Chemical Sensing

<i>Asim Kumar Ray</i>	1241
54.1 Analyte Requirements	1242
54.2 Brief Review of Inorganic Materials	1243
54.3 Macrocylic Compounds for Sensing	1245
54.3.1 Preparation of Sensing Membranes	1245
54.3.2 Thin-Film Properties	1248
54.4 Sensing with Phthalocyanine and Porphyrin	1250
54.4.1 Amperometric Sensor	1250
54.4.2 Optical Sensors	1251
54.4.3 Detection of Volatile Organic Vapour Compounds	1254
54.5 Polymeric Materials	1255
54.5.1 Conducting Polymers	1255
54.5.2 Ion Sensing	1257
54.5.3 Examples of Other Polymeric Sensors	1257
54.6 Cavitand Molecules	1259
54.7 Concluding Remarks	1261
References	1262

55 Packaging Materials

<i>Darrel Frear</i>	1267
55.1 Package Applications	1268
55.2 The Materials Challenge of Electronic Packaging	1269
55.2.1 Materials Issues in High-Speed Digital Packaging	1270
55.2.2 RF Packaging Materials Issues	1271
55.3 Materials Coefficient of Thermal Expansion	1272
55.4 Wirebond Materials	1272
55.4.1 Wirebonds for Digital Applications	1272
55.4.2 Wirebonds for RF	1273
55.5 Solder Interconnects	1273
55.5.1 Flip-Chip Interconnects	1276
55.5.2 Flip Chip for RF	1277
55.5.3 Pb-Free	1277
55.6 Substrates	1278
55.6.1 RF Substrate Materials	1279
55.7 Underfill and Encapsulants	1280
55.7.1 Underfill	1280
55.7.2 Encapsulation	1280
55.8 Electrically Conductive Adhesives (ECAs)	1281
55.8.1 Adhesive Polymers	1281
55.8.2 Metal Fillers	1282
55.8.3 Conduction Mechanisms	1282
55.8.4 Isotropic Versus Anisotropic Conduction	1282
55.8.5 Rework	1283
55.9 Thermal Issues	1283
55.9.1 Thermal Issues in Digital Packaging	1283
55.9.2 Thermal Issues in RF Packaging	1284

55.10 Summary 1284

References 1285

Acknowledgements 1287

About the Authors 1291

Detailed Contents 1307

Glossary of Defining Terms 1333

Subject Index 1367

Glossary of Defining Terms

π Band

Equivalent of the valence band in an organic semiconductive material

σ Bond

Bond formed along the internuclear axis, between the nuclei of the atoms. In ethylene, a σ bond is formed by the overlap of two carbon sp^2 hybrid orbitals

π Bond

Chemical bond formed between the nuclei of atoms. There are two bonding regions above and below a nodal plane containing the nuclei. In ethylene, a π -bond is formed by the overlap of p orbitals on the carbon atoms

π^* Band

Equivalent of the conduction band in an organic semiconductive material

A

Abbe Number

(v_d) a measure of the dispersion (wavelength dependence) of the refractive index n of transparent materials at visible wavelengths.

$v_d = (n_d - 1)/(n_F - n_C)$, where $d = 587$ nm, $F = 486$ nm, and $C = 656$ nm. Typically v_d takes values of 80–20, which decreases with increasing n_d from 1.5 to 1.9

Absorption Coefficient

(α) The absorption coefficient represents the logarithmic decrement of the incident light intensity per unit length in the material

Acceptor

Impurity in a semiconductor or any other electroluminescent device capable of inducing hole conduction and accepting a valence-band electron to produce an acceptor energy level

Acceptor Level

A level with energy E_a that acquires electron captured from the valence band to an orbit near the acceptor's atom

Acceptor Concentration

Number of acceptors cm^{-3} . It refers to the total concentration N_a and the concentration of ionized acceptors N_a^-

Accumulation

The condition in which, for an MIS structure, the applied voltage to the gate electrode V_G results in an increase in the density of majority carriers near the surface of the underlying semiconductor. The applied voltage is less than the flatband voltage, $V_G < V_{FB}$

ACRT

Accelerated crucible rotation technique: a means to stir liquids in sealed ampoules, leading to improvements in crystal properties

Active Material

The main functional component of the thick-film paste, usually in the form of a finely divided powder

Active Matrix Addressing

The technique used to write and change images on complex liquid-crystal displays. Each display consists of thousands of picture elements (pixels) arranged as a matrix of columns and rows. Each pixel has a transistor switch [thin-film transistor (TFT)], which has been deposited onto one plate of the liquid-crystal display. A voltage is applied sequentially to the rows of pixels, and selected pixels within each row are switched on by data voltages on the columns. As the next row in the sequence is activated, the elements of the previous row are switched off from the data line by the TFT, but hold their charge as an electrical capacitor until the row is addressed next time in the sequence, after a full frame time. Thus the image is built up line by line, and individual pixels must hold their charge for the period of the full frame-address time

Air Mass Number

This is often abbreviated to AM followed by a number and characterizes the intensity and spectrum of light (modified by absorption in the atmosphere). The AM number is 0 for no air mass (i. e. for a satellite solar array), 1 for the sun directly overhead and 1.5 for the sun at 45° to the horizon

All-Optical Switch

Optical inputs are switched by other optical signals. The switching is produced in many cases by refractive-index changes. Comparing electronic devices such as transistors, in which an electrical input is controlled by an electrical current, we expect that the all-optical switch can work faster, being free from electrical noises. However, at present, in contrast to the μm scale of a transistor, an all-optical switch may have cm scales

Amorphous Metaloxides

Glassy alloys of a transition metal with oxygen, typical examples being TiO_2 , Ta_2O_5 , Nb_2O_5 , and Y_2O_3 . In bulk form, these materials are typically polycrystalline or crystalline ceramics. However, amorphous thin films can be deposited with relative ease, and they have been widely used as high-index layers in optical filter design and as dielectric layers in the microelectronics industry

Amorphous Network

Network consisting of atoms distributed randomly with a short-range order, i. e., holding a regular coordination of atoms

Amorphous Semiconductors

Semiconducting materials having a random network without long-range order

Amorphous Silicon

Thin films of silicon with a structure that is noncrystalline but which have the advantage of a higher absorption coefficient compared to crystalline silicon

Amphiphilic

An organic molecule possessing both polar (hydrophilic) and nonpolar (hydrophobic) chemical groups. The classic examples are fatty acids, which consist of a nonpolar ($-\text{CH}_2-$) chain terminating in a polar $-\text{COOH}$ acid group

Amphoteric Dopants

Dopants that may serve both as acceptors and donors

Amphotericity

This occurs when an impurity can act either as an acceptor or a donor. In GaAs, a Si atom can take up a site on either sublattice and can act as a shallow donor or shallow acceptor

Analytical Electron Microscopy

The application of transmission and/or scanning electron microscopy for the purpose of structural and/or chemical microanalysis

Anion

The electronegative atomic component in a compound, e.g. As in GaAs, Se in ZnSe

Annealing

Low-temperature heat treatment, normally in the presence of Hg, with the aim of adjusting the electrical properties of materials. The process of heating and slowly cooling semiconductor materials to reverse lattice damage. As a metallurgical term annealing describes the use of heat to reduce the internal energy of a crystal. This energy may be mechanical (strain) or may represent variations in concentration, which can be reduced by diffusion, etc.

Antiferromagnetically Coupled Media (AFM)

See synthetic ferrimagnetic media (SFM)

Anti-site Defect

Occurs in semiconductor compounds and is an anion (cation) replacing a cation (anion) on a regular cation (anion) lattice site e.g. As_{Ga} (Ga_{As}). The anion As_{Ga} (cation Ga_{As}) anti-site is expected to be a donor (acceptor) in the III–Vs and similarly in the II–VIs

Apparent Bandgap Narrowing

Definition of apparent bandgap seen in bipolar transistors due to the combined effects of bandgap narrowing due to heavy doping and the use of Fermi–Dirac statistics

Areal Density

Number of bits per unit area stored on a recording medium. In disk drives, it is the product of the bits/unit length times the number of tracks per unit length (usually inches and the areal density is frequently quoted in Gbits/in²)

Arrhenius Relation

A general expression connecting the value of a parameter in a thermally activated process to the absolute temperature T . So a diffusivity $D = D_0 \exp(-Q/k_B T)$ where D_0 , Q and k_B are the pre-exponential factor, the activation energy for diffusion and Boltzmann's constant, respectively

Astigmatism

A defect in the performance of a lens arising from the lens asymmetry about the optic axis, leading to loss of sharpness of an image at focus

A-Swirl

A network of dislocations loops. The origin is a stacking fault of Si interstitials. The size of the dislocation loops can be many μm

Attenuated Total Reflection (ATR)

Technique to measure infrared absorption spectra of thin films on a prism with enhanced sensitivity utilizing multiple reflections of an infrared beam in the prism

Auger Electron Spectroscopy (AES)

A method in which an electron spectrometer is used to measure the energy distribution of Auger electrons emitted from a surface

Auto-compensation

Often associated with amphotericity, auto-compensation occurs when the incorporation of electrically active impurities of one type leads to the automatic incorporation of electrically active defects of the opposite type. As an example, the doping of melt-grown GaAs with donors always leads to inadvertent incorporation of acceptors at about one tenth of the donor concentration

B

Back End of Line (BEOL)

A term generally describing the final points in the IC fabrication process line, such as the interconnection and packaging fabrication steps. These process steps generally require lower thermal budgets

Backscattered Electrons

Incident electrons that have interacted with atomic nuclei and scattered backwards with comparable incident energy. They may be used for sample imaging with contrast dependent on the local averaged atomic number

Baldereschi Concept

Mean-value k -points in the Brillouin zone are defined such that the value that any given periodic function of

the wavevector assumes at these special k -points is an excellent approximation to the average value of the same function throughout the Brillouin zone

Band Offset

The valence- and conduction-band offsets describe the corresponding band-edge discontinuities at intimate and abrupt semiconductor–semiconductor interfaces

Band Alignment

Type I: electrons and holes are confined within the same layer. Type II: electrons and holes are confined in different (adjacent) layers

Band Filling

An effect that takes place when a semiconductor is illuminated with light at a frequency within the absorbing region. Upon absorption of the incident photons, the electrons undergo a transition from an occupied band (VB) to the empty band (CB), thereby partially populating the empty band

Bandgap (BG), Band Gap or Energy Gap

The forbidden energy gap between the valence band, normally filled with electrons, and the conduction band, normally empty of electrons. The bandgap energy measures the energy difference between the top of the valence band and the bottom of the conduction band. The *optical BG* is the photon energy above which semiconductor absorbs energy of incident EM radiation. It may be direct or indirect, depending on type of electron transition from VB to CB the radiation induces; in Si, the optical BG is indirect with $E_g = 1.17$ eV at 4.2 K

Bandgap Engineering

The science of tailoring the semiconductor band gap to control the electrical and optical properties of the material

Barrier Height

The barrier height of a Schottky contact is the energy distance right at the metal–semiconductor interface from the Fermi level to the majority-carrier band-edge, i. e., the conduction-band maximum in n-type and the valence-band maximum in p-type contacts

Beam Effective Pressure Ratio

In molecular beam epitaxy, this is the ratio between the partial pressures of the various components of the molecular source beams

BEEEM (Ballistic-Electron-Emission Spectroscopy) Technique

A method to determine the barrier heights of Schottky contacts by tunnel injection of almost monoenergetic electrons from a biased metal tip into metal–semiconductor contacts that pass the metal films as ballistic electrons and are collected by the semiconductors, provided their energy is high enough to overcome the interfacial barriers

Birefringent Crystals

Crystals such as calcite that are optically anisotropic, which leads to an incident light beam becoming

separated into ordinary (o-) and extraordinary (e-) waves with orthogonal polarizations; incident light becomes doubly refracted because these two waves experience different refractive indices n_o and n_e

BLIP

Abbreviation for background-limited performance, which describes the best signal-to-noise performance that can be achieved with a detector. In this condition the only noise stems from the random arrival rate of photons and is therefore at a theoretical minimum

Blocking Temperature

Temperature where the magnetic exchange coupling between an antiferromagnet and a ferromagnet vanishes

Boltzmann Equation (BE)

Generally an integro-differential equation that governs the carriers' distribution function. At equilibrium the BE is automatically satisfied by a Fermi distribution. The generalized BE is a version revised to include quantum effects in the carrier's scattering and interaction with EM fields. Methods for the solution of the BE include the relaxation time, variational, displaced Maxwellian and Monte Carlo methods

Boule

There are many terms used for large crystals. These include boule, ingot and crystal. The top of the crystal is called the seed end, the first to solidify, etc. The bottom is called the tail or tang end

Bragg's Law

Can be summarized by the defining equation $\lambda = 2d_{hkl} \sin \theta$, which must be satisfied for constructive interference to occur, corresponding to maxima in the positions of diffracted intensity

Bravais Lattice

A repetitive arrangement of points in space such that the environment of each point is identically similar to that of every other point. 14 such arrangements are used to describe crystal systems

Bremsstrahlung

Electromagnetic radiation produced by the rapid deceleration of an electron in the vicinity of an atomic nucleus

Bridgman Growth (Technique)

Technique developed in order to grow compounds where one component is volatile at the growth temperature. The growth is carried out by physically withdrawing the melt from its furnace, while the melt crystallizes on a seed. The volatile component is replenished from a reservoir in a separate furnace. The arrangement of the equipment can be horizontal or vertical, the former offering the possibility of viewing the top of the solid/liquid interface and giving visual feedback of the growth process

Brillouin Zone (BZ)

A primitive cell in the reciprocal-space lattice, which proves to be body-centered cubic (BCC) for Si. The

BZ provides the domain of wavevectors k , which relate to the electron momentum in crystal $p = \hbar k$.

Broadband Sensitizer

Typically a species that is co-doped along with rare-earth ions into a glass host to increase the pumping efficiency or radiative efficiency of the rare-earth ions. Various sensitizers have been demonstrated, including silicon nanoclusters, silver ions, and other rare-earth ions (such as in the sensitization of erbium by ytterbium)

Brooks–Herring (BH) Formula

Celebrated equation for the carrier's momentum relaxation time due to scattering by ionized impurities

Bulk Modulus

Ratio of the compressive or tensile force applied to a substance per unit surface area to the change in volume of the substance per unit volume

C

C/V Technique (Capacitance–Voltage Technique)

The C/V technique is a method to determine barrier heights of Schottky contacts from the variation of their capacitance as a function of applied voltage

Calamitic

Denotes the rod-like shape of the molecules forming the majority of liquid-crystal phases that, therefore, are known as calamitic liquid crystals

Capacitance Equivalent Thickness (CET)

Theoretical thickness of SiO₂ that would be required to achieve the same capacitance density as an alternative dielectric material. This is extracted directly from, for example, the accumulation region of a C–V curve and has no further correction

Capacitance–Voltage (C–V)

C–V measurements of a MOS capacitor can be used to determine the oxide thickness and the amount of electrically active defects (fixed charge, trapped charge, mobile charge, and interface trapped charge) and mobile contaminants in the oxide. It can also be used to determine the oxide thickness and the profile of dopants at the semiconductor surface. It is a common characterization technique used in wafer fabrication facilities

Carrier Mobility

Measure of the effect of applying an electric field to charge carriers (electrons or holes). It is the additional velocity acquired by the carriers (drift velocity) divided by the electric field

Carrier Range

Determines how far an injected carrier moves per unit field before becoming immobilized in a deep trap. Sometimes, it is calculated as the product of the drift mobility and lifetime

Carrier Concentration

Number of the current carriers per cm^{−3}. It can be

electron concentration n and/or holes concentration p . The intrinsic carrier concentration (n_i) is $n = p$ in a perfect crystal semiconductor that has not been doped (i. e., an intrinsic semiconductor). The value of n_i for Si is approximately to 1.08×10^{10} cm^{−3} at 300 K

Carrier Velocity, Microscopic

Gradient of specific band spectrum $E_s(k)$ with respect to p

Cascaded Second-Order Materials

Materials that exhibit second-order nonlinear properties that act as an effective third-order nonlinear material

Cathodoluminescence

Luminescence stimulated by energetic electrons

Cation

Electropositive atomic component in a compound, e.g. Ga in GaAs, Cd in CdS

Cd-Based Compound Semiconductor

Semiconductor that contains the element Cd

Centro-symmetric

A crystal structure in which a center of inversion exists is called centro-symmetric. There are 32 crystal classes, among which 11 are centro-symmetric

Cermets

Standard form of thick films requiring exposure to a high-temperature firing cycle, typically in the range 700–1000 °C

Chalcogenide Glasses

Glasses that are amorphous alloys containing S, Se, and/or Te. Typical examples include Se, GeS₂, GeSe₂, As₂S₃, As₂Se₃, and As₂Te₃. By intermixing these and other binary chalcogenide glasses, a wide variety of multicomponent glasses can be formed. Further, a wide range of nonstoichiometric compositions is possible. Several compositions have become standard industrial materials, including Ge₃₃As₁₂Se₅₅ and Ge₂₈Sb₁₂Se₆₀. The chalcogenide glasses are characterized by narrow bandgaps and good transparency in the mid- to far-infrared wavelength range

Characteristic X-Ray

An X-ray photon generated by the transition of an outer-shell electron to an inner-shell vacancy of an atom, with energy characteristic of the element

Charge-Blocking Layer

A layer that is sometimes placed between the conductive layer and the subsequently coated layer (usually the *charge-generation layer*) to prevent charge injection in the dark

Charge Carriers

Generally current-carrying (free) electrons and holes in conduction and valence bands, respectively, and electrons trapped by nearby impurity atoms

Charge Generation

Conversion of absorbed light into charges that are injected into, and subsequently transported through,

the *charge-transport layer* of an organic photoconductor

Charge-Generation Layer (CGL)

A layer in an organic photoconductor in which light is absorbed and converted into electrically charged species

Charge-Generation Material (CGM)

The light-absorbing material in the *charge-generation layer*

Charge Pumping

Electrical technique used to determine the interface state density in MOSFETs. In its simplest implementation, voltage pulses of well-defined properties of amplitude, rise and fall times and period are applied to the gate terminal of a MOSFET with the source and drain terminals grounded, and the substrate current is measured as the base level of the gate pulse is swept from below accumulation to above threshold voltage. The resulting substrate current is then used to extract the interface trap charge density (cm^{-3})

Charge Transport

involves electron hopping through a polymer matrix containing charge-transport moieties (see MDP). The charge carrier for hole transport is the radical cation and that for electron transport is the radical anion of the charge-transport moiety

Charge-Transport Layer (CTL)

The layer in an organic photoconductor through which charges generated in the CGL are transported

Charge-Transport Material (CTM)

The *charge-transport layer* dopant that transports charge injected from the *charge-generation material*

Chemical Annealing

A concept to explain crystal formation by the assistance of chemical species such as atomic hydrogen during film growth

Chemical Vapor Deposition (CVD)

Technique for the growth of thin solid films on a crystalline substrate as the result of thermochemical vapor-phase reactions. This reaction occurs above a solid surface, e.g. a diamond single crystal, which causes deposition onto that surface. In the CVD technique the thin films are obtained under equilibrium conditions. The gas-phase reaction is activated by temperature in order to create condensable species that lead to film growth. Different CVD techniques have been developed (See: *Mercury-sensitized photo-CVD*, *Hot-wire CVD*, *MOCVD*, *MPCVD*, *PECVD*). All techniques used for the production of diamond require a means of activating gas-phase carbon-containing precursor molecules, usually CH_4 , diluted by hydrogen

Chiral Nematic

Liquid-crystal phase in which the director is twisted into a helical arrangement; the phase is, therefore, not

superimposable on its mirror image and so exhibits macroscopic chirality. The chirality of the phase originates from that of the constituent molecules. The chiral nematic is sometimes referred to as a cholesteric phase

Chromatic Aberration

A blurring of resolution due to the differing focal lengths of a lens when acting on electrons of differing energy

Cleave and Stain

This is a rapid, though destructive, technique for assessing semiconductor structures. Particular chemical mixtures affect the surfaces of different semiconductor types in different ways, giving rise to effects that can be seen in a simple optical microscope. The sample is first cleaved to produce a clean surface, preferably at an angle that tends to magnify the scale of the structure, and then stained

CMT

A narrow-bandgap II–VI semiconductor compound, $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, which is still the preeminent infrared (IR) material

Coefficient of Thermal Expansion (CTE)

Increase in the length of a macroscopic material for a given temperature increase. The units of CTE are $1/^\circ\text{C}$

Coercive Field

The electric field required to reduce the polarization in a ferroelectric material to zero in a fully saturated ferroelectric hysteresis loop

Coercive Squareness Parameter (S^*)

The slope of the major hysteresis curve of a ferromagnetic recording layer normalized to the ratio of the remanent magnetization to the coercive field $S^* = 1 - [M_r/H_c / dM/dH(H = -H_c)]$

Coherence Length

ξ is the shortest distance within which a considerable change of the Cooper-pair density is possible

Cohesive Energy

The energy required to dissociate the atoms of a solid into isolated atomic constituents

Columnar Phase

A liquid-crystal phase in which the constituent molecules, usually disc-like, are stacked into columns; these are arranged parallel to each other in either a hexagonal or rectangular array

Complementary Metal Oxide Semiconductor (CMOS)

An MOS device technology consisting of nMOS and pMOS transistor pairs

Complex

A bound state between two or more adjacent, like or unlike species, e.g. an anti-site and a lattice vacancy at a nearest-neighbor site on the same sublattice such as $\text{As}_{\text{Ga}}\text{V}_{\text{Ga}}$

Compositional Uniformity

A critical parameter in all ternary alloy systems that

determines a material's usefulness in device applications

Compound Semiconductor

Semiconductor crystals composed of two or more atomic elements from different groups of the periodic chart

Compressive Strain

Type of strain obtained when a strained $\text{Si}_{1-x}\text{Ge}_x$ layer is grown on a silicon substrate

Compton Scattering

Transfer of energy from a photon to an electron, leading to the scattering of a longer-wavelength photon

Concentration Quenching

the reduction in luminescence efficiency and luminescence lifetime of a laser glass when the rare-earth dopant concentration is high. Quenching is due to interactions between closely spaced rare-earth ions at high concentrations. These interactions create new pathways, other than the desired radiative decay, for the ions to relax to the ground state after they have been raised to a desired lasing level by pump energy

Conduction Band (CB)

Band of energies allowed for electrons in a semiconductor, which is empty in undoped semiconductors. There are many CBs, separated by band gaps. However, CB term usually refers to the conduction band that has the lowest energies, or closest to the valence band (VB). It is separated from the VB by a band gap

Conductive Layer

The layer on which the *charge-generation layer* and *charge-transport layer* are coated. It is connected to ground in the electrophotographic process

Conductivity Mass

Used for electron or hole conductivity (mobility) calculation

Conductivity

Product of the number of electrons or holes per unit volume, the electronic charge and the mobility of the carrier $\sigma = ne\mu_{e,h}$

Configuration Coordinate Model

Model of interaction between electronic system of a luminescent center and the vibrational system of the surrounding atoms and ions

Converse Piezoelectric Effect

The generation of mechanical strain in a material in response to an applied electric field, where the strain is linearly proportional to the applied field

Conversion Efficiency

Ratio of incident photon energy and output electricity energy

Cooper Pairs

Bound electron pairs formed of electrons with opposite spins and momenta (in their ground state with zero electric field)

COP

A crystal-originated particle (COP) is a small aggregate or cluster of vacancies. Usually, it is of octahedral shape with (111) facets, but in the case of nitrogen doping, it can also adopt the form of a platelet or a rod-like cluster. The usual size is 70–200 nm

Corona Discharge

When a high voltage of 5–10 kV is applied to sharp edges or fine wires of metals, the gas such as air surrounding the metal is ionized and discharged. If the voltage is DC, ions having the same polarity with the DC voltage move toward a grounded electrode

Correlation Factor (f)

The jumps of a diffusing particle are generally not random but correlated, i. e. successive jumps are not independent of each other; f (always ≤ 1) is the ratio of the jump rate of correlated jumps to the jump rate if the jumps were random

Cotterell Atmosphere

The stress field around a dislocation can be reduced if defects congregate around them. This results in defect atmospheres, called Cotterell atmospheres, which are commonly observed even when the dislocation itself would not provide an observable signature. EL2 defects aggregate in this way in SI substrate material

Coupled Granular Continuous (CGC) Recording Media

In perpendicular recording media: a recording surface consisting of two layers: the first being granular and the second consisting of multilayers of typically cobalt and palladium

Critical (Transition) Temperature

The T_c is a characteristic property of the superconductor in question. At T_c the resistance drops abruptly to an unmeasurably small value

Critical Field

B_c is the magnetic field sufficient to destroy the superconducting state in a *type I superconductor*

Critical Fluctuations

Variations of the superconducting order parameter $\psi = n_C^{1/2} \exp(i\varphi)$ (where n_C is the Cooper-pair density) around the thermal equilibrium values

Critical Thickness

Maximum thickness of a strained layer that can be grown on a substrate before relaxation of the layer occurs

Cryopanel

Liquid-nitrogen-cooled surface within the growth chamber arranged to minimize contamination of the wafer surface by impurities

Current In the Plane (CIP)

Magnetoresistive device with the sense current flowing in the plane of the films

Current Perpendicular to the Plane (CPP)

Magnetoresistive device with the sense current flowing perpendicular to the plane of the films

CZ Crystals

Crystals grown according to the so-called *Czochralski technique*

Czochralski and Liquid-Encapsulated Czochralski (LEC) growth

Czochralski growth *pulls* a crystal from the melt. The seed is held in a holder vertically above the melt and rotated. After partial immersion into the melt, it is slowly withdrawn and the crystal forms on the seed. The growth takes place because of the increased heat loss from the crystal compared to the melt. Rotation of the seed and growing crystal leads naturally to a near-cylindrical crystal from which circular substrates can be sawn. The GaAs melt must be protected against As loss if this method is to be used. Boric oxide (see liquid encapsulation) liquifies on the surface of the melt and, to some extent, on the crystal and acts as a barrier against As loss. The boric oxide also insulates the melt from the crucible; reducing the rates at which impurities from the crucible are introduced into the melt and aiding withdrawal of the exhausted melt from the crucible after Growth

D**Dangling Bonds (Broken Bonds)**

Point-like defects that are produced by breaking covalent bonds. The dangling bond may have an unpaired electron, two paired electrons, or no electrons. These dangling bonds were denoted as D^0 , D^- , and D^+ by Mott, the superscript representing the charge state

Dark Decay

Rate of decrease of the surface potential of a charged organic photoconductor

Data Rate

Rate at which the digital data, recorded or read back in a digital storage device, is transferred. Characterized by Mbytes/s or MB/s, where a byte (B) is eight bits

Debye Length

A measure of the distance beyond which the electric field due to a point charge (e.g. an ionized dopant atom) is screened by the free carriers (electrons or holes) and becomes increasingly negligible

Deep-Energy-Level Impurities

Doping impurities or other impurities whose energy level lies toward the center of the bandgap

Deep-Level Transient Spectroscopy (DLTS)

DLTS is a useful technique for measuring concentrations of deep levels in conducting semiconductors. Deep levels within the depletion layer of a pn junction are filled by putting the junction into forward bias. On applying a reverse bias, these

levels are allowed to empty thermally. This produces a change in junction capacitance consisting of a series of exponentials. From the temperature spectra generated for different time windows of the transient, Arrhenius plots can be made to determine the energy level (eV) and capture cross section (cm^2) of the defects

Defect Chemistry

Representation of interactions between defect species and free carriers (dopant, impurity, native defect, electrons and holes) in terms of chemical-style equations from which the concentrations of the various species concerned can be obtained

Deformation Potential

Effective electric potential describing interaction of carriers with the lattice deformation irrespective of what is the cause of the deformation. It describes the phenomenology of the interaction between carriers and phonons of deformation types

Deformation Potential Parameters

There are two and four such parameters for electrons and holes in Si, respectively. For holes the parameter set includes one describing interaction with anti-phase deformation of atoms in the nonequivalent primitive cell's positions

Density of States

Widely abbreviated as DOS, the density of states is the number of energy states per unit energy interval at the energy E . It stands for the density distribution of allowed electronic energies in a material and is widely used with respect to this distribution across the bandgap of disordered semiconductors

Density-of-States Mass

Used for electron or hole density-of-states calculations

Depletion

The condition in which, for a MIS structure, the applied voltage to the gate V_G electrode results in a reduction of majority carriers near the surface of the underlying semiconductor. This region is referred to as the depletion layer. Charge remaining in this near-surface region is due to ionized dopants present in the channel region. This condition occurs for applied voltages between the flatband and threshold voltages, i. e. $V_{FB} < V_G < V_{thr}$

Depth of Field

The distance along an optic axis that an object can be moved without noticeably reducing the resolution

Depth of Focus

Maximum spacing between an imaging screen and a photographic plate (or CCD) that allows a recorded image to be retained in focus

Depth Profiling

Monitoring of signal intensity as a function of a variable that can be related to distance normal to the surface, cf. compositional depth profile. Note: the

signal intensity is usually measured as a function of sputtering time

Depth Resolution

Depth range over which a signal changes by a specified quantity when reconstructing the profile of an ideally sharp interface between two media or a delta layer in one medium

Detection Efficiency

Percentage of radiation incident on a detector system that is actually detected

Detective Quantum Efficiency (DQE)

The ratio of the square of the signal-to-noise ratio at the output of the detector to that at the input. The relative increase in image noise due to an imaging system as a function of the *spatial frequency*, f , is expressed quantitatively as $DQE(f)$, which represents the signal-to-noise transfer efficiency for different frequencies of information in an image

Detectivity

A signal-to-noise parameter, normalized to area and bandwidth, often used for photoconductive detectors or single-element photovoltaic detectors

Devitrification

The transition of a glassy material to its lower-energy crystalline state. This process is usually driven by thermal energy, such as if the material is held at some characteristic temperature above its glass-transition temperature. The difference between the crystallization temperature and the glass-transition temperature for a particular glass is one measure of its stability

DFT (Density-Functional Theory)

DFT is a quantum-mechanical approach to many-body electronic structure calculations of molecular and condensed-matter systems. The many-electron wave-function is written in terms of the electron density. The major problem with DFT is that the exact functionals for exchange and correlation are not known. The widespread approximations are the local-density approximation (LDA) and GW approximation (GWA). The LDA assumes the functionals for exchange and correlation to depend only on the density at the coordinate where the functional is evaluated. LDA DFT calculations dramatically underestimate the gaps of sp^3 -bonded semiconductors and insulators. In the GW approximation, the exchange and correlation is expressed as the product of a single-particle propagator G and a screened interaction W . GWA DFT calculations yield the gaps of sp^3 -bonded semiconductors and insulators generally to within a few tenths of an eV

Diamond Structure

A structure in which each atom lies at the center of a tetrahedron surrounded by four nearest neighbors located at the points of the tetrahedron. In such a structure diffusion is isotropic

Dielectric Constant/Susceptibility (DC/DS)

Basic material optical tensor property that linearly connects the electric displacement to the electric-field vector inside the material. For Si it is scalar. At optical frequency DC/DS is a complex quantity

Dielectric Materials

A class of materials that are insulators or nonconductors where charge imparted to one part of the material is not communicated to any other part

Diffraction

The process of scattering and reconstruction of radiation in specific directions as a consequence of interaction with a periodic structure, e.g. light interacting with a grating, or X-rays interacting with a crystal lattice

Diffusion Coefficient or Diffusivity

The macroscopic parameter that characterizes the jump rate or jump frequency of a species at the atomic level. It is normally obtained by matching experimental profiles to solutions of the appropriate diffusion equation, e.g. Fick's second law. The diffusion coefficient measurement techniques include Haynes–Shockley, time-of-flight and noise-measurement-based methods

Diffusion Length

Measure of the spatial extent of a diffusion region. Typically the concentration of a diffusant can be expected to fall by more than a factor of 10^3 over three diffusion lengths

Digital Printing

Technologies in which the image is comprised of a pattern of pixels. In electrophotographic digital printing the electrostatic latent image is written on a charged photoreceptor using a computer-controlled laser or LED bar

Dipping

Liquid-phase epitaxy technique in which the substrate is lowered into the melt in a vertical orientation

Direct Piezoelectric Effect

The generation of an electric polarization change or a charge separation in a material in response to an applied stress, where the polarization change is linearly proportional to the applied stress

Director

Symmetry axis for properties such as the refractive index or dielectric tensor of a liquid-crystal phase; at a molecular level it is commonly associated with the preferred orientation of the unique axis of the constituent molecules, either rod-like or disc-like

Disclotic

Indicates the disc-like shape of the anisotropic molecules that constitute a class of liquid-crystal phases known as disclotic liquid crystals

Disorder Model

Derived from Monte Carlo simulation studies, this

model describes electronic transport in random media in terms of disorder-induced fluctuations of both hopping site energy and relative orientation

Distribution Function

A function of p (impulse) and the carrier's coordinate, which obeys *Boltzmann's equation* and presents a semiclassical probability description of a carrier's statistics and dynamics. At equilibrium, it converts to the Fermi function over the carrier's energy, which in turn reduces to a Maxwell-Boltzmann distribution for lightly doped semiconductors

Donor

An impurity in a material that is capable of inducing electrical conduction in that material by transferring an electron to the conduction band. A donor's level is the energy E_d of an electron in an orbit near the donor atom. The donor concentration is the number of donors cm^{-3} , referring to either the total concentration N_d and the concentration of ionized donors N_d^+

Dopant

An atom from a different group of the periodic table from the host atom that substitutes for it. It can be a donor, acceptor or amphoteric and is deliberately incorporated to give n- or p-type conductivity

Doped Semiconductor

A semiconductor that contains donors and/or acceptors. *Lightly* and *moderately doped* semiconductors contain impurities with energy levels that are well separated from the CB and VB. *Heavily doped* semiconductors (n^+ - or p^+ -type) contain so many dopants that their energy levels merge with CB or/and VB, which leads to metallic conduction at low temperatures. For n-Si, the critical donor concentration of this, Mott's insulator-metal transition, proves to be about $3 \times 10^{18} \text{ cm}^{-3}$. Fermi-Dirac statistics must be used for carriers in n^+ - or p^+ -type semiconductor, but in n^+ - or p^+ -Si full degeneracy of carriers is met only well below room temperature

DRAM

Dynamic random-access memory

Drift Mobility

Velocity per unit of applied field imparted to a charge carrier by the electric field. In the presence of carrier traps it is reduced from the free-carrier velocity by the fraction of time a carrier spends in the trapping centers

Drude Formula

Phenomenological equation for the free-carrier contribution to the *dielectric constant/susceptibility*. It contains two parameters: the effective plasma frequency and the relaxation time

Drum Photoreceptors

Devices where the layers are coated sequentially on a metal drum substrate

Dual-Layer Organic Photoconductor

A photoconductor with an architecture where the

charge-generation and charge-transport functions are carried out in separate layers (see *charge-generation layer* and *charge-transport layer*, respectively)

Dynamic Coercivity

The change of coercivity of small ferromagnetic particles with switching time under thermal excitation

E

E' Center

A kind of unpaired-electron dangling bond in oxide glasses. In SiO_2 , an Si atom that is bonded to three O atoms may have one E' center. The center, which may be produced by radiation, gives ESR signals and optical absorption at $\hbar\omega \approx 6 \text{ eV}$

E Layer

Thin ferromagnetic cobalt alloy film added to one or two sides of the Ru layer in a synthetic ferrimagnetic media to increase the exchange coupling

Easy Axis

Direction in a ferromagnetic sample along which the magnetization is oriented in the absence of an external magnetic field. In thin films, the easy axis is the direction in a substrate surface along which the director tends to align; it is determined by the nature of the surface treatment

Edge-Defined Film-Growth Technique

Shaped crystals, including tubes, sheets, etc., are grown through a die placed on the melt surface

Effective Mass

Generally a set of parameters describing the dynamics of the current carriers, which may deviate drastically from the free-electron mass. For Si, it includes the effective mass tensor of electron in the CB (m_t , m_l , m_l) and the masses of heavy (m_1), light (m_2) of spin orbital split-off (m_3) holes in the VB. Generally, $m_{1,2}$ depend on the direction of hole's momentum.

Different averages over this direction define masses that enter the conductivity and the DOS

Einstein Relation

An equation connecting the mobility μ , (speed per unit electric field) of a charged particle to its diffusivity D . Specifically $\mu = qD/kT$, where q is the electric charge on the particle. This expression is only valid for nondegenerate material

EL(2)

Electron level number 2. The deep levels in most semiconductors were labeled according to their observation during DLTS measurements. EL levels are donors (they are neutral when they have their electron). HL levels are acceptors (they are neutral when they have their hole). EL2 is the most important deep donor level (actually, a double donor) and is either the As anti-site defect or contains the As anti-site as a component

Elastic Compliance Constant (C)

Defined from Hooke's law by $X = Ce$, where X is the stress and e is the strain. It has units of pressure

Elastic Deformation

Deformation of a body in which the applied stress is small enough that the object retains its original dimensions once the stress is released

Elastic Stiffness Constant (S)

Defined from Hooke's law by $e = SX$, where e is the strain and X is the stress. It has units of inverse pressure

Elasticity

A property of liquid crystals that causes the directors to be uniformly aligned parallel to each other. Deviations from this uniform ground state require the addition of elastic energy to the liquid crystal

Electrically Conductive Adhesives (ECA)

Composite materials consisting of a dielectric curable polymer and metallic conductive particles. The polymer is an adhesive material that chemically reacts with metals and other polymers to form a bond. The metallic particles in the ECA form a network in the cured joint that form an electrical conduction path

Electro-absorption

Generic term for all effects of changing the absorption coefficient upon applying strong electric field. For Si, the dominant is the change of the free carriers' absorption in IR

Electrochemical Potential

In semiconductors (*see Fermi level*)

Electron Affinity

The energy distance from the conduction-band minimum to the vacuum level at the semiconductor surface

Electron-Beam Lithography

Method for micro- and nanoscale fabrication where a pattern on a polymeric layer (resist) is exposed to electron irradiation

Electron

Elementary particle having a negative charge of 1.602×10^{-19} C and rest mass m_0 equal to 9.109×10^{-31} kg

Electron-Phonon Scattering

In Si, in addition to scattering by deformation phonons, scattering may also occur with short-wavelength (inter-valley) phonons

Electro-optic Effect

A change in refractive index upon application of a strong electric field. This is linked to electro-absorption through the Kramers–Kronig relation. It manifests itself as a change in *birefringence* in response to an applied electric field

Electrophotography

Printing technology in which charged marking particles are developed on an organic photoreceptor with an image pattern of surface charge, the

electrostatic latent image, and the subsequent transfer of these particles to a receiver

Electroplating

Process used to deposit a material on a conducting substrate in a chemical bath containing metal ions using electrical current

Eley–Rideal Reaction

A surface-reaction scheme in which an arriving chemical species react on the surface without any surface diffusion process. It is considered the counterpart of a Langmuir–Hinshelwood reaction

Ellipsometry

Method of defining the optical constants by measuring the ratio of reflectance for s- and p-polarized radiation, and the relative phase shift between the two, at large incidence angles

Emission Device

An emission device creates photons from extra electron–hole pairs through a process called radiative recombination. The electron–hole pairs are often created by forward-biasing a pn junction

Encapsulation

A dielectric polymer material that flows over and covers the electronic components in a package to provide mechanical and electrical protection to the packaged devices

Energy Gap

see Bandgap

Energy-Band Spectrum

Dispersion relation $E_s(\mathbf{k})$ of allowed electron energy to \mathbf{k} within the *Brillouin zone*. As a function of small deviation of \mathbf{k} from the extrema points of $E_s(\mathbf{k})$ it proves to be anisotropic parabolic for the conduction band and non-parabolic for valence band, forming ellipsoid valleys and warped spheres, respectively, in the *Brillouin zone*

Epitaxial Layer (or Epilayer)

Thin crystalline layer on a single-crystal substrate with orientation and lattice structure determined by the substrate crystallography

Epitaxy

(Greek; arrange upon). This is the growth of one crystal (the epitaxial layer) on the surface of another (the substrate) and where the orientation of the grown layer is the same as the substrate. If the layer is of the same material as the substrate, the growth is homoepitaxial. If the layer and substrate are different, the growth is heteroepitaxial. Usually, epitaxy can only be performed where there is close match between the lattice constants, although it is possible for a layer of quite different lattice constant to be grown if there is a change in crystal orientation between the layer and the substrate. The term epitaxy was introduced to describe the importance of having parallelism between two lattice planes with similar networks of closely similar spacing

Equivalent Oxide Thickness (EOT, t_{eq})

The theoretical thickness of SiO₂ that would be required to achieve the same capacitance density as an alternative dielectric material. This is often determined by quantum-mechanical modeling of the C - V curve for thin dielectric layers (where tunnelling effects result in substantial “leakage” current) and/or where poly-Si gate electrodes (resulting in depletion effects) are present

Etch Pit Density (EPD)

This is the usual way in which substrate suppliers quote dislocation densities. EPD is found by etching the surface in molten alkalis, like KOH, and viewing the resulting pitted surface under an optical microscope. Each pit represents a single dislocation ending at the surface

Ewald Sphere

Geometrical construction used to illustrate the relationship between a diffraction pattern and the reciprocal lattice of a diffracting crystal

Exchange Field (H_{ua})

Parameter characterizing the strength of the coupling of a ferromagnetic to an antiferromagnetic film. Measured in Oe or A/m

Exciton

An electron-hole pair, which behaves like a H atom, with a binding energy of 10–100 meV and lifetime of ≈ 1 ns. The wave-function, in principle, extends over the whole crystal

Exhaustion Concentration

$N = |N_d N_a|$ is the maximum attainable carrier concentration in a doped semiconductor

F**Fermi Level or Fermi Energy**

For a metal, the Fermi level is defined as that energy below which its one-electron levels are occupied and above which they are empty in the ground state, i. e., at zero temperature. In the context of semiconductors, the term Fermi level is a synonym for their *electrochemical potential*

Ferroelastic

A material that exhibits a phase transition from a non-strained to a strained state, generating a spontaneous strain at the transition, and in which the spontaneous strain can be switched between two or more stable states by the application of a stress

Ferroelectric Domains

Adjacent regions in a ferroelectric crystal that have their spontaneous polarization vectors inclined to one another

Ferroelectric Hysteresis

The loop that is produced when the polarization in a ferroelectric material is plotted as a function of applied electric field

Ferroelectric Relaxor

A ferroelectric that shows a broad peak in relative permittivity at the paraelectric-to-ferroelectric phase transition and in which the temperature of the peak is strongly dependent upon measuring frequency

Ferroelectric

A polar dielectric in which the polarization can be switched between two or more stable states by the application of an electric field

Fick's First Law of Diffusion

For diffusion parallel to the x -axis the diffusant flux (atoms per unit area per second) is equal to $-D\partial c/\partial x$ where c is the diffusant concentration which defines the diffusivity D

Fick's Second Law of Diffusion

Based on the first law, this law gives the rate of build up of the diffusant concentration at a given depth as $\partial c/\partial t = \partial(D\partial c/\partial x)/\partial x$

Field-Effect Transistor (FET)

A transistor where the current between two electrodes (the drain and source) is modulated by the electric field from a third electrode (the gate)

Field Emission

Electron emission from a metal or semiconductor into vacuum under the influence of a strong electric field

Figure-of-Merit (FOM)

Measure of the performance of an integrated circuit technology. The time delay associated with signal propagation is a common metric

Fill Factor (FF)

Measure of the maximum power that can be obtained from a photovoltaic solar cell compared with the product $I_{sc} \cdot V_{oc}$

Firing

One of the key stages of the thick-film production process. It is usually undertaken in a continuous-belt furnace at temperatures of up to 1000 °C

Flatband Voltage (V_{FB})

The voltage applied across an MIS device at which there exists no charge in the semiconductor. As a result, the valence and conduction band structure of the semiconductor is flat. This condition occurs when this voltage equals the workfunction difference between the metal electrode and the semiconductor under ideal conditions. The presence of charges in the insulator or at the interface, due to defects, modifies the voltage required to achieve the flatband condition

Flexoelectricity

Generation of a macroscopic electrical polarization in a nematic liquid crystal when the director distribution is deformed from its uniform state of alignment. The inverse of this effect is the deformation of the director distribution when an electric field is applied to a nematic; the magnitude of the deformation is linear in the field

Flip Chip

Integrated-circuit-level interconnect that can be used to replace wirebond interconnects. The flip chip is a solder bump on an area array on the chip surface that routes the power, ground and signals from the integrated circuit to the bumps. The metallization on the IC surface is called the under-bump metallurgy that forms a metallurgical bond with the solder. The flip-chip die is joined to the package by placing the die face down on the matching bond pads on a substrate and reflowing the solder to form an electrical, thermal and mechanical interconnect

Floating Gate Current

Very small gate currents (\approx fA or less) can be measured using the floating gate technique in which the drain current of a MOS transistor is measured after the gate bias has been removed. Then, by using the decay in the drain current as a function of time and the measured drain current versus gate voltage characteristics and the oxide capacitance, the gate current can be calculated

Floating Zone (FZ) Technique

This technique uses a solid feed rod that is melted at its lower end by a high-frequency coil. The melt flows through a central hole of the coil down to the growing crystal below the coil. The FZ technique does not need a crucible. Hence, the melt is not contaminated by other materials. The crystals that are grown according to the FZ technique are called FZ crystals

Fluence

This is the total, time-integrated, flux of particles (electrons, protons, ions, etc.) that reach a unit area of sample. It can be used to represent the total number of ions implanted into a surface and is sometimes called the dose

Fluorescence

Luminescence with a lifetime \lesssim 10 ns

Fluoride Glasses

Multicomponent glasses, typically based on fluorides of zirconium, barium, lead, gallium, lanthanum, aluminum, and sodium. They have a wide transparency range, from ultraviolet to mid-infrared wavelengths. They also have low characteristic phonon energies and can dissolve large concentrations of rare-earth ions. For these reasons, they are extremely popular as hosts for rare-earth-doped amplifiers and lasers operating in the UV-visible and mid-infrared regions

Flux Lines (Vortices)

Regions in which magnetic flux enters a type II superconductor in the mixed state. Screening currents flow around each of the flux lines. Due to the repulsive vortex-vortex interactions a hexagonal flux line lattice is formed

Flux Pinning

An effect caused by defects. The energy for the

formation of the normal cores of the flux lines is reduced in regions with a reduced Cooper-pair density

Flying Height

Spacing between the bottom of the slider and the top of the recording medium in a disk drive

Forced Convection

Process of forced flow of fluid

Fourier-Transform Infrared (FTIR) Spectrometry

A means of performing infrared absorption measurements with great speed and precision, based on an optical interferometer

Four-Point Probe

A popular technique used to measure the resistivity (Ω cm) of a semiconductor sample. Four collinear probes are equally spaced and a current is applied through the outermost probes and the voltage difference across the two inner probes is measured. From the voltage and current values, the resistivity can be calculated assuming that the contact and spreading resistances between the probes and the semiconductor and the resistance of the probes are negligible in comparison to the resistance of the semiconductor

Franck-Condon Principle

Absorption and emission of light takes place in so short a time that the atomic coordinates in ground and excited states are unchanged; i. e. transitions are vertical on a configuration coordinate model

Free Convection

The process of movement of fluid through a temperature gradient

Free Film

Ferromagnetic film in a magnetoresistive or spin-valve read head in which the magnetization can respond easily to an external magnetic field

Friedel's Law

A law that states that the intensity of an hkl reflection in a diffraction pattern is equal to the intensity in the opposite $\bar{h}\bar{k}\bar{l}$ reflection

Front End Of Line (FEOL)

A term generally describing the initial points in the integrated-circuit fabrication process line, such as the transistor fabrication steps. These processes generally require higher thermal budgets

G**Gate Oxide Integrity (GOI)**

The GOI measures the breakdown stability of the oxide layer of a MOS capacitor when a voltage is applied to the capacitor. There are different types of gate oxide integrity tests. Most of them are standardized. This may currently be the best method to detect the density of very small voids

Geminate Recombination

The primary recombination of a correlated

(quasi-bound) hole–electron pair immediately following photoexcitation

Ghosting

The change of X-ray sensitivity of the X-ray image detector as a result of previous exposure to radiation. In the presence of ghosting, a shadow impression of a previously acquired image is visible in subsequent uniform exposures

Giant-Magnetoresistive (GMR) Effect

Change in the resistivity of a stack of ferromagnetic films coupled by thin non-ferromagnetic films when the orientation of the magnetization of the films adjacent to the non-ferromagnetic film is varied. The change in resistivity with low temperatures and high magnetic fields was observed to be as large as 50%; hence the adjective *giant*

Glass Ceramics

Glass that contains crystalline particles or regions. Such materials may be transparent or smoggy

Glass Frit

Binding matrix within the thick film. This binds the active particles together and also bonds the thick film to the substrate

Glass-Transition Temperature

The approximate temperature at which a material changes from a supercooled liquid to an amorphous solid, or vice versa. The transition is marked by an abrupt but continuous change in slope of the specific volume and enthalpy versus temperature curves. Viscosity varies rapidly near the glass-transition temperature, which is also sometimes called the softening temperature

Glow Discharge Spectrometry (GDS)

A method in which a spectrometer is used to measure relevant intensities emitted from a glow discharge generated at a surface. This is a general term that encompasses glow discharge optical-emission spectrometry (GDOES) and glow discharge mass spectrometry (GDMS). GDOES is a method in which an optical-emission spectrometer is used to measure the wavelength and intensity of light emitted from a glow discharge generated at a surface. GDMS is a method in which a mass spectrometer is used to measure the mass-to-charge quotient and abundance of ions from a glow discharge generated at a surface

Gradient Freeze Growth

This technique is similar to the *Bridgman method* except that the melt is not physically removed from the furnace. Instead, the temperature gradient along the melt is controlled electrically so that solidification commences at the seed and progresses until the melt is exhausted. Once again, both horizontal and vertical arrangements are possible and often the reservoir is replaced by liquid encapsulation to impede As loss

from the melt. This is now the favored technique for the growth of most GaAs substrate material

Grading

Grading is the gradual change from one semiconductor to another, exemplified by the gradual change from GaAs to GaAsP that is necessary in some LEDs

Gratzel Cell

Named after its inventor, Michael Gratzel, this is a dye-sensitized cell using a porous TiO₂ substrate to collect the photo-generated charge

GRINSCH Laser

A graded-refractive-index separate-confinement heterojunction laser (*see SCH Laser*). Optical confinement is optimized by grading the composition of the cladding layers

Gunn Diode

(*See negative differential resistance*). This is a device based on low-doped n-type GaAs and relies on the NDR effect. It is used for the generation of low-power microwave currents

H

Hall Effect

The deflection of a charged particle moving in a magnetic field that is perpendicular to its motion. The deflection is due to the Lorentz force on the charged particle and it causes the charges to accumulate in one side of the sample. The voltage measured at right angles to the current flow is called the Hall voltage. The Hall effect can be used to characterize the mobility (cm²/Vs), resistivity (Ω cm), type of carrier and carrier density (cm^{−3}) of a semiconductor sample. The *Hall mobility* is the product of conductivity and the Hall constant (the transverse-electric-field Hall field divided by the product of the current density and the magnetic induction) for a conductor or semiconductor; a measure of the mobility of the electrons or holes in a semiconductor. The *Hall coefficient* and *factor* relate μ^H linearly to σ and μ , respectively

Hard Axis

Direction in a ferromagnetic sample at right angles to the easy axis

Hard Magnetic Bias Film (also Longitudinal Bias Film)

Permanent magnetic film abutted to the free film in a magnetoresistive or spin-valve head to eliminate magnetic domains

Head Field Slope Parameter (Q)

The maximum slope of the magnetic field of a write head in the direction of the magnetization of the recording layer normalized to the coercive field (H_c) divided by the magnetic spacing (d), $Q = \frac{dH_x/dx}{H_c/d}$

Heat Capacity

The amount of heat required to change the

temperature of a substance temperature by one degree, with units of energy per degree

Heteroepitaxy

The growth of a layer of markedly different composition from the substrate, i.e. the epitaxial layer and the substrate are made from different materials, e.g. the diamond growth on an iridium single-crystal substrate. Another example is the growth of $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$ layers lattice-matched to InP. The heteroepitaxial growth techniques are chemical-vapor deposition, liquid-phase epitaxy, and molecular-beam epitaxy

Heterojunction Bipolar Transistor (HBT)

A modification to the standard bipolar transistor where a heterojunction is used to control the carrier flow at the base

Heterojunctions and Heterostructures

A heterojunction is the junction between different materials (e.g. GaAs and AlGaAs with different bandgaps). Such a junction exhibits several properties that may be very useful to device manufacturers. These include changes in the energies of the valence or conduction bands, or both, and changes in the optical properties. A heterostructure is a semiconductor structure where the properties of heterojunctions are exploited

Heterojunction Laser

A semiconductor laser where both electrical and optical confinement exploit the properties of heterojunctions is a heterojunction laser. If the active region is sandwiched between two heterojunctions, the laser is termed a double heterojunction

Hexagonal As

The natural form of elemental arsenic

HgMnTe

An alternative Hg-based ternary system to *CMT*

High-Electron-Mobility Transistor (HEMT)

This is a modification of the *metal semiconductor field-effect transistor (MESFET)* where the separation of the electrons in the channel from the ionized donors is achieved by using a heterojunction. This results in higher electron mobility, and thus greater speed and lower noise

High Index Contrast

Waveguides or devices fabricated using two or more materials that have very different refractive index. High index contrast is the basis for the confinement of light to very small cross-sectional-area waveguides or very-small-volume optical cavities, either using total internal reflection or photonic bandgap effects. High index contrast is thus the basis for increased density of optical integrated circuits

High-Temperature Solution Growth

Slow cooling, or top-seeded (similar to Czochralski) growth of materials in a solution designed to reduce

the freezing point of the desired phase below a critical temperature

Homeotropic Alignment

Alignment of the director perpendicular to the surface of a substrate

HOMO – Highest Occupied Molecular Orbital

The highest energy molecular orbital of an atom or molecule that contains an electron. If the atom or molecule were to lose an electron, it would most likely lose it from this orbital

Homoeptitaxy

Epitaxial growth of a layer of the same composition as the substrate

Homojunction

Junction between materials of the same chemical type but of different electrical properties. For example, a p–n junction in GaAs is a homojunction

Hopping Conduction

Electronic transport that localized electrons hop from site to site with the assistance of phonons

Hot-Carrier Phenomena

Group of effects associated with the carrier transport in high electric fields. For Si, it includes the nonlinear relation between drift velocity and electric field with saturation (n-Si) and near to saturation (p-Si), anisotropy of drift velocity regarding the electric orientation relative to the crystallographic axes (the Sasaki–Shibuya effect) and the diffusion's anisotropy as regards the density gradients along and perpendicular to the strong electric field

Hot-Wire Chemical Vapor Deposition (CVD)

The same technique as catalytic CVD, which is used for thin-film growth through decomposition of source gas materials utilizing catalytic reaction on the heated filament

Huygens's Principle

Concept used to explain the propagation of a wave (and diffraction processes), in that every point on a primary wavefront acts as a source of spherical wavelets, the envelope function of which acts to reconstruct the primary wavefront a short time later

Hybrid Circuits

Circuits consisting of electronic elements made from differing enabling technologies such as thick-film, thin-film, monolithic silicon, etc.

Hybridization

In molecules, the orbitals occupied by the electrons are seldom pure s or pure p orbitals. Instead they are mixed or hybrids, formed by combining the s and p orbitals. Examples are sp (linear), sp² (trigonal and planar) and sp³ (tetragonal)

Hydrothermal Growth

Growth that takes place in aqueous-based solutions at high temperatures and high pressures

Hysteresis

The state of the magnetization in response to an external magnetic field depends on the initial state of the magnetization

***I/V* Technique (Current–Voltage Technique)**

The *I/V* technique is a method to determine barrier heights and ideality factors of Schottky contacts from their current–voltage characteristics

IDCA

Abbreviation for for integrated detector cooler assembly, a commonly used infrared detector scheme in which the detector is mounted directly on the cold finger of a cryocooler (often based on Stirling cycle engines). The detector and cold finger are then enclosed in a vacuum vessel with a transparent window and optical baffles

Ideality Factor

The ideality factor of Schottky contacts characterizes the variation of their barrier heights as a function of applied voltage

Image Lag

Lag is the carryover of image charge generated by previous X-ray exposures into subsequent image frames. The residual signal fractions following a pulsed X-ray irradiation are referred to as image lag

Impact Ionization

Process of generation of electron–hole pairs by carriers moving in an electric field higher than the breakdown threshold, which for Si is of the order of 3×10^5 V/cm

Improper (or Extrinsic) Ferroelectric

A ferroelectric material in which the spontaneous polarization is not the primary order parameter. Frequently, the primary order parameter is the spontaneous strain associated with a ferroelastic phase transition

Impurity

A foreign atom unintentionally present in the semiconductor, incorporated either during growth or processing

***In Situ* Monitoring**

Tools to help in understanding layer growth kinetics and to provide monitors suitable for feedback control in epitaxial growth systems

Inelastic Scattering

The result of a collision between a photon or electron and the nuclei or electrons of a material, such that there is a net change in the internal energies of the system and in the sum of the kinetic energies before and after the collision

Infrared-Laser-Absorption Spectroscopy (IRLAS)

Technique to determine density of gas-phase

molecules and chemical species using infrared absorption spectra

Infrared-Reflection-Absorption Spectroscopy (IRRAS)

Technique to determine the surface-bonding configuration on the film-growing surface using light absorption during reflection of infrared beam

Integrated Circuit (IC)

Combination of active and passive circuit elements to enable computational logic or analog operations

Integrated Optics/Photonics

The manufacture of photonic elements and circuits on a planar substrate, typically using thin-film deposition, lithography, and etching steps. Typically, the substrate is a glass or semiconductor wafer and the photonic elements are guided-wave devices

Interconnect

The system of metal conducting lines and contacts among IC components. The categories include local (between neighboring devices), intermediate (between neighboring circuit elements), and global (across the IC chip)

Interdiffused Multilayer Process (IMP)

The process of obtaining a uniform alloy composition of $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ by the growth of alternate layers of the binary compounds that are thin enough to completely interdiffuse within the time of growth

Interface Trap Density

The density (cm^{-3}) of positive or negative charges located at the silicon/silicon dioxide interface, due to defects induced by oxidation, structural defects, impurities or other defects caused by bond-breaking mechanisms such as radiation or hot carriers. These states are typically in electrical communication with the charges in the channel of a MOSFET and results in stretch-out of the capacitance–voltage characteristics. They also affect the turn-on and turn-off characteristics of a MOS transistor

Interface-Induced Gap States

Because of the quantum-mechanical tunnel effect, the wave-functions of electrons tail across semiconductor interfaces in energy regions where occupied states overlap a bandgap. These evanescent waves are the continuum of the *intrinsic* interface-induced gap states

Intergrain j_c

Macroscopic transport critical current density, which can be much smaller than the intragrain j_c because of the weak-link behavior of large-angle grain boundaries in the cuprate superconductors

Intermetallic Compounds

Substances that form between pad metallizations and the active components of molten solder (typically Sn). For Cu metallization, the Sn reacts to form Cu_3Sn and Cu_6Sn_5 intermetallics. For Ni, the Sn reacts to form Ni_3Sn_4 . These intermetallics typically form at the interface of the solder and the metallization and

are usually more brittle than the solder or the metallization

Intermetallic Reactions And Phases

Two or more metals can react chemically to produce compounds. These are revealed in phase diagrams as labeled, vertical lines. The compound is often referred to as an intermetallic phase

Interstitial

A site lying between regular lattice sites that can be occupied by dopant, impurity or host atoms. The latter case is known as a self-interstitial. They generally behave as a donor

Inter-subband (ISB) Transition

Transition between confined states within the conduction or valence bands

Intragrain j_c

is the value of j_c within a single grain, limited only by the pinning properties

Intrinsic Point Defect

Is the general term for either a vacancy or a Si interstitial in the Si matrix

Inversion

Condition in which, for a MIS structure, the applied voltage to the gate electrode V_G results in an increase of minority carriers near the surface of the underlying semiconductor. This region is referred to as an inversion layer. When present in a MISFET, this condition results in a conducting channel between the source and drain regions. This condition occurs for applied voltages beyond the threshold voltage, i. e. $V_{th} < V_G$

Inversion Symmetry

A system in which the laws of physics are unchanged by the operation of inversion

Ion Implantation

This is a method of introducing impurities into the surface of a solid. The impurity atoms to be introduced are ionized and accelerated by a high voltage, up to 10^6 V in some instances. They penetrate the surface to a depth dependent on their energy. Unfortunately, this disrupts the crystal lattice by introducing irradiation damage and this must be repaired by annealing at elevated temperatures before the electrical activity of the implanted atoms can be obtained

Ionization

Process of forming an electrically charged atom (ion)

Ionization Energy

The energy distance from the valence-band maximum to the vacuum level at the semiconductor surfaces

IPEYS technique (Internal Photoemission Yield**Spectroscopy)**

IPEYS is a method to determine barrier heights of Schottky contacts and band offsets of semiconductor heterostructures by photoinjection of hot electrons over the energy barriers at metal/semiconductor and

semiconductor/semiconductor interfaces, respectively, as a function of the photon energy of the exciting light

Irreversibility Line

Situated well below the upper critical field, this line separates a region without pinning and zero j_c , above the irreversibility field, from a region with pinned vortices ($j_c > 0$)

Irvin Curve

The experimental resistivity ρ of Si at room temperature versus N , the free-carrier concentration

K

Kikuchi Lines

Diffuse background of lines in a diffraction pattern formed by the elastic scattering of incoherently scattered electrons

Knudsen Cell

High-vacuum oven from which molecular beams can be produced in molecular-beam epitaxy. These ovens are usually fitted with temperature sensors. Shutters mounted in front allow the beams to be turned on and off, as required

Kramers–Kronig Relations (KKR)

Dual integral relations between the real and imaginary parts of one-sided Fourier transform of causal functions of time, which includes all physical admittances. The KKR allow us to define the optical constants based on KKR relations for amplitude and phase of complex reflection coefficient from opaque sample of material under measurement

Kypopoulos Technique

This is similar to the Czochralski method but the extent of pulling is limited, producing crystals of greater diameter but shorter length

L

Langmuir–Blodgett Film Deposition

Method to build up multilayer structures of organic materials by the transfer of a monolayer floating on a water surface to a solid substrate as the latter is raised and lowered through the monolayer/water interface. Deposition modes include Y-type (monolayer transfer from the water surface to the substrate on both the upward and downward motion of the latter through the monolayer/air interface), X-type (film transfer only on downward motion of the substrate) and Z-type (film transfer only on the upstroke)

Laporte's Rule

Transitions between states of the same parity are forbidden

Laser Crystallization

A useful method to fabricate polycrystalline semiconductor showing high electron mobility by

irradiating laser beam on amorphous semiconductor thin films

Laser-Induced Fluorescence (LIF)

Technique to determine the density of gas-phase atoms and molecules whose optically excited states are emissive states

Lattice-Matching

Growth of quantum wells on a substrate with the same, or very similar, lattice constant

Law of Mass Action

An equation linking the concentrations of the various reacting species in a chemical-type equation, derived from defect chemistry analysis

Lifetime

Minority-carrier lifetime is one of the electrical parameters that determines the performance of IR devices; it can be affected by impurities and/or dopants

Light-Induced Defect Creation

The defect creation associated with illumination of samples. This is normally observed in amorphous materials having a flexible network

Lineage

A reduction in total dislocation strain energy can be obtained either by polygonization or by dislocations arranging themselves in linear arrays. The latter results in the creation of a small-angle grain boundary or lineage. The reader will find examples of this in most books on materials science

Liquid Crystal

State of matter with properties characteristic of both a liquid and a crystal; that is, it flows like a liquid and certain properties are anisotropic like those of a crystal. At the molecular level the phase has long-range orientational order and some element of translational disorder at long range

Liquid-Encapsulated Czochralski Technique

An inert layer, usually B_2O_3 , is used, which floats on the top of the melt to prevent loss of volatile components; this is used for As- and P-containing compounds

Liquid-Phase Epitaxy (LPE)

LPE is epitaxial growth of a layer by dissolving the required material in a liquid. On cooling, the material becomes supersaturated and forms a solid film. If this film is deposited on a substrate the growth can be epitaxial

Localized Vibrational Mode (LVM) Absorption

Form of mid- to far-infrared absorption resulting from simple vibrations of light atoms in a heavier lattice. Such vibrations do not couple well to lattice vibrations and result in sharp absorption lines whose frequencies are directly related to mass and whose strengths are proportional to concentration. LVM absorption is particularly suited for measuring carbon concentrations in SI GaAs

Longitudinal Magnetic Recording

Recording system consisting of write and read head and a recording media where the magnetization is parallel to the surface of the recording medium

Long-Range Disorder

Defined by the absence of the long-range order, i. e. the lack of periodic arrangements of atoms

Low-Dimensional Structure

Heterostructure with dimensions comparable to the wavelength of an electron or hole, typically less than 100 Å, so that quantum effects are important

Low-Frequency Noise

Electrical measurement of the current or voltage noise spectral densities of a semiconductor component, typically at frequencies from ≈ 1 Hz to ≈ 100 kHz. The units are in A^2/Hz for current noise spectral density and V^2/Hz for voltage noise spectral density. It is a sensitive electrical technique that can be used to probe microscopic defects in semiconductor components and as a gauge of their reliability

Low-Temperature Co-fired Ceramic (LTCC)

A composite material structure made of alumina bonded at a temperature below the sintering temperature of alumina using a glass binder. In a substrate form, LTCC can be co-fired (melting the glass and bonding the alumina while capturing the metal interconnects and lines) at lower temperatures so it is possible to use higher conductivity materials like Cu or Ag conductors in the ceramic

Low-Temperature Solution Growth

Normally used for water-soluble materials, with growth being progressed by either slow cooling or solvent evaporation

LPE

Liquid-phase epitaxy is growth of a (normally) thin epitaxial layer from the liquid phase onto a suitably lattice-matched substrate by dipping, tipping or sliding boat means

L-pit

see A-swirl

Luminescence

In general terms, is the emission of light by a luminescent material (also called phosphor) due to conversion of a certain type of energy into electromagnetic radiation over and above thermal radiation. The luminescence is the light emitted by nonthermal sources in contrast with the emission of radiation from a heated object, which is called incandescence. In accordance with the source of energy the luminescence may be the photoluminescence (excited by external illumination), electroluminescence (induced by the passage of electrical current), cathodoluminescence (excited by irradiation with electrons), triboluminescence (excited by mechanical treatment, e.g. grinding), chemiluminescence (emitted during chemical

reaction), bioluminescence (appearing as a result of biological processes), thermoluminescence (caused by the rise of temperature) and so on

LUMO – Lowest Unoccupied Molecular Orbital

Refers to the lowest-energy molecular orbital of an atom or molecule that does not contain an electron. If the atom or molecule were to accept an electron, it would be most likely to do it with this orbital

LW

Abbreviation for long wave, this term refers to the transparent atmospheric window between the wavelengths 7.5 μm and 14 μm (sometimes called the thermal band). The edges of this window are influenced by the water-vapor content and atmospheric conditions

M

Magnetic Annealing

Thermal process involving a ferromagnetic sample in a magnetic field in which the sample has induced a direction of easy magnetization (easy axis)

Magnetic Random-Access Memory (MRAM)

Digital memory device composed of ferromagnetic thin films coupled to current-carrying conductors used to establish the direction of the magnetization in the ferromagnetic films. The resistance of the coupled films is different when the direction of magnetization in the two ferromagnetic films is changed

Magnetic Spacing

Spacing between the bottom of the slider and the center of the ferromagnetic storage layer in a recording medium characterized by the magnetic spacing parameter d

Magnetic Transition

In digital magnetic recording, the region between opposite states of the magnetization in the recording medium. The length of the transition (l) is $l = \pi a$, where a is the transition parameter

Magnetic Tunneling Junction (MTJ)

MRAM in which the magnetic state of the ferromagnetic sample is sensed by the change in resistance of electron current tunneling through a thin insulating layer when the magnetization in the two electrodes coupled to the ferromagnetic sample is changed from parallel to antiparallel orientation

Magnetoresistance

Change in resistance of a ferromagnetic material due to changes in the orientation of the magnetization with respect to an induced easy axis. Sometimes referred to as the anisotropic magnetoresistance (AMR). Characterized by the change in resistivity with magnetic field normalized to the nominal resistivity $\Delta\rho/\rho$

Magnetoresistivity (MR)

Change of resistivity ρ in applied magnetic field. For Si, MR is a diverse tensor reflecting the symmetry of the CB and VB

Majority Carriers

Electrons in an n-type and holes in a p-type semiconductor. Majority-carrier mobility is the mobility of electrons (holes) in n-type (p-type) material

Media Flux

Product of the remanent magnetization and the thickness of a magnetic recording film

Media Noise

Noise voltage at the terminals of a read head caused by fluctuations in the magnetization of the recording media

Mercury-Sensitized Photo-CVD

Technique for preparing thin films by decomposing source gas materials by collision with photo-excited mercury atoms

Mesogen

This is a molecular compound or mixture of compounds capable of forming a liquid-crystal phase over a range of temperatures between the crystal and isotropic liquid phases

Metal Semiconductor Field-Effect Transistor (MESFET)

Probably the simplest three-terminal device that can be fabricated on GaAs. The “metal semiconductor” refers to the gate structure, which is a simple Schottky barrier. The application of a reverse bias to this gate produces a depleted region that occludes part of the cross-sectional area of the channel, and thus modifies its conductivity

Metallic Nanocomposites

Composite materials made out of metallic quantum dots embedded in organic or glass hosts

Metastability

A term that refers to the nonequilibrium nature of glasses or amorphous solids. Amorphous solids have excess internal energy relative to the corresponding crystalline state or states of the same material. The method of manufacture, such as melt quenching, inhibits a transition to the lowest-energy crystalline state

Microphotonics

Refers to the chip-scale manufacture of optical and photonic waveguide circuitry, using processing techniques borrowed from the microelectronics industry. Related to this is the need for high-density integrated optics, as facilitated by high-index-contrast waveguides and photonic crystals. By usual definition, microphotonics refers specifically to the monolithic manufacture of optical and photonic elements on silicon (CMOS) chips

Miller's Rule

An empirical rule proposed by Miller in 1964, which suggests that $\chi^{(2)}/\{\chi^{(1)}\}^3$ is nearly constant for all non-centrosymmetric crystals

Miniband

Interval of allowed energies for carriers in a superlattice, resulting from the delocalization and broadening of the quantum-well energy levels

Minority Carriers

Electrons in a p-type and holes in an n-type semiconductor. Many electron devices work on the base of minority-carriers transport. Minority-carrier mobility is the mobility of electrons (holes) in p-type (n-type) material

MIS

Abbreviation for structure metal/insulator/semiconductor (see MOS)

Misfit Dislocations

These are dislocations introduced near the boundary of an epitaxial layer and the substrate, when a mismatch between the lattice constants exists. The density of these dislocations is proportional to the mismatch. Although they run parallel to the boundary, they can interact and penetrate much of the epitaxial layer; see threading dislocations

Mixed State

In type II superconductors this state exists between the lower and upper critical fields. In the mixed state superconducting and normal regions coexist. Magnetic flux enters the superconductor in the normal cores of the flux lines

Mobility Edge

The boundary between localized and delocalized states in a band

Mobility Gap

Energy separation between two mobility edges of the conduction and valence bands

Mobility

Parameter defined as the ratio of the carrier velocity (cm/s) to the electric field through which it is moving (V/cm). It is expressed in units of cm^2/Vs . At the microscopic level, it is related to the dominant scattering time and the effective mass of the carrier. The notion includes drift mobility μ , which corresponds to the motion parallel to the field and Hall mobility μ^H , which corresponds to motion perpendicular to the electric field, when a magnetic field is applied

MOCVD

Technique for depositing thin films of compound semiconductors from a chemically reactive vapor phase where at least one of the components is a metal-organic

Modulation Transfer Function (MTF)

The MTF measures the efficiency of an imaging system such as a detector to resolve (transfer) different

spatial frequencies of information in an image. In other words, the MTF is the relative signal response of the system as a function spatial frequency

Molecular-Beam Epitaxy (MBE)

Low-temperature growth technique for epitaxial films from atomic or molecular beams from thermal evaporation sources. It is carried out in ultra-high vacuum. CBE is a form of molecular-beam epitaxy in which group III and group V beams are generated from external gas sources. MOMBE is a form of MBE where some or all beams are generated from external gas sources, in III–V MBE only the group III beams, from metal-organic sources. GSMBE is a form of MBE where some or all beams are generated from external gas sources, in III–V MBE only the group V beams, from hydride sources

Molecular Electronics

Exploitation of organic materials for electronics and optoelectronics applications. Examples are displays based on liquid crystals and organic electroluminescent polymers. In the case of organic polymers, the area is sometimes referred to as plastic electronics

Molecular Reorientation

Nonlinear process in which the orientation of molecules in liquid changes upon illumination with intense light

Molecularly Doped Polymers (MDP)

Formulations of charge-transport materials dissolved in a polymer matrix. CTLs of OPCs are typically MDPs

Monolayer (ML)

A single layer of atoms (in III–V MBE a layer of the compound or alloy, e.g. a layer of Ga + As)

Moore's Law

prediction based upon the empirical observation by G. Moore that the minimum cost of manufacturing integrated circuits per component actually decreases with the increase in the number of IC components, and thus with greater circuit functionality and computing power. A corollary of this observation is that the density of integrated circuit transistors will double roughly every 1.5–2 years (see scaling)

MOS

Abbreviation for a metal/oxide/semiconductor structure, which is one possible realization of MIS. The insulator has traditionally been SiO_2 (the oxide) due to superior materials and electrical properties, resulting in the abbreviation MOS

Moseley's Law

States that the square root of the frequency of characteristic X-rays, for certain elements, is linearly related to the atomic number

MOVPE

Metalorganic vapor-phase epitaxy is a low-temperature growth technique using metalalkyls

(and elemental Hg in the case of Hg-based ternary compounds) as the sources

MPCVD Technique

(Microwave plasma chemical deposition technique) A technique for film deposition. This is now the most widely used technique for diamond growth. Energy is transferred by the microwaves to gas-phase electrons, which transfer their energy to the gas through collisions. The gas molecules dissociate and the active species are formed and the deposition of diamond onto the substrate immersed in the plasma occurs

Multi-Crystalline Silicon (mc-Si)

Wafers of silicon, cheaper to produce than the single-crystal silicon wafers but have multiple single-crystal grains in each wafer

Multiple Quantum Well (MQW)

Structure composed of a stack of QWs separated by sufficiently thick barriers so that the electron wave functions and energy levels are localized to each QW

MW

Abbreviation for medium wave, the term for the transparent atmospheric window between the wavelengths 3.3 μm and 5 μm . There is a CO_2 absorption band around 4.25 μm which divides the band into two with the better atmospheric transmission in the 3.3 to 4.2 μm band

N

Narrow Bandgap

Refers to a semiconductor with a forbidden energy gap of less than about 0.7 eV, making it suitable for detection in the infrared wavebands

Native Defect

A vacancy, a self-interstitial, an anti-site or any complex of these

Negative Differential Resistance (NDR)

(see, transferred electron effect). In GaAs, electrons that undergo the transferred electron effect, are excited from the primary conduction band minimum where their effective mass is low to subsidiary minima where their effective mass is considerably greater. As a result, although they have greater kinetic energy, their drift velocity is lower. This, in turn, results in a reduction of current in the external circuit. The current-voltage characteristics show a reduction in current after a critical voltage is applied (although the ratio of current to voltage will always be positive). This is NDR. It can be shown that current flow in the NDR region is unstable and that charge tends to be transported in groups, resulting in oscillations in the current flow. This is the Gunn effect and is used in some microwave sources

Nematic Phase

is the simplest of the liquid-crystal phases; it has

long-range orientational order but is devoid of long-range translational order

NETD

Noise-equivalent temperature difference is a measure of sensitivity for a multiplexed infrared detector and is the change in scene temperature that produces a signal equivalent to the rms noise level of the detector

Nonlinear Directional Coupler

A device where a part of one of two waveguides is made out of a nonlinear material. Changing the intensity of the incident light changes the effective path length experienced by the waveguided light, thus controlling coupling between the two waveguides

Nonlinear Fabry-Perot Interferometer

A device that consists of two mirrors separated by a nonlinear material. As the refractive index of the nonlinear material changes with an increased level of illumination, the effective path length of the resonator is altered, changing the transmission properties of the device

Nonlinear Figure Of Merit

A figure of merit that describes the applicability of a refractive nonlinear optical material in terms of nonlinear index change and effective absorption

Nonlinear Kramers-Kronig Transformations

A set of two relations that permits calculation of the real (imaginary) part of nonlinear response at a given optical frequency given the knowledge of the imaginary (real) nonlinear response at all other optical frequencies

Nonlinear Mach-Zehnder Modulator

A Mach-Zehnder modulator in which a part of the waveguide is made out of a nonlinear material. Changing the intensity of the incident light changes the effective path length experienced by the light in the nonlinear waveguide, controlling transmission

Nonlinear Material Figures Of Merit

A set of metrics that allows us to quantify the applicability of a nonlinear optical material to optical switching

Nonlinear Optical Medium

A material whose properties change upon illumination with intense light

Nonlinear Optical Switching

Switching of optical signals using an illumination-dependent phase shift of nonlinear materials

Nonlinear Periodic Structure

A Bragg periodic structure in which at least one set of layers is nonlinear

Nonlinear Phase Shift

A difference between the phase shift experienced by intense light and a phase shift experienced by light which intensity approaches zero

Nonlinear Polarization

Polarization that does not experience purely linear dependence on the electric field

Non-stoichiometry

Defined as the difference between the total anion and cation concentrations, hence anion- or cation-rich

n-Type Conductivity

A semiconductor material, with electrons as the majority charge carriers, that is formed by doping with donor atoms

Nucleation Temperature

Temperature where stable defect clusters/aggregates start to form as nuclei

Numerical Aperture

Figure of merit used to describe the power of a lens, depending on the angle of collection of the lens and the refractive index of the medium in which the lens is situated

O

Onsager Model

A model that describes the probability of geminate recombination as a function of applied field with the efficiency of geminate pair photogeneration and the initial pair-separation distance as parameters

Optical Absorption Coefficient

4π times extinction index, divided by the radiation wavelength

Optical Bistability

A phenomenon in which the instantaneous transmittance of the device depends both on the level of incident illumination and on the prior transmittance of the device. Such an element enables all-optical switches

Optical Constants

The real and imaginary parts of the square root of the *dielectric constant/susceptibility*, called the refraction and extinction index, respectively; they define the optical properties of bulk material and mesoscale structures

Optical Emission Spectroscopy (OES)

A technique to determine the plasma parameter and reactions in the plasma by measuring the emission intensity of line spectra from the plasma

Optical Gap

(See *Bandgap*)

Optical Limiter

A device in which the transmittance decreases with increased level of illumination

Optical Properties of Sample

These include the apparent reflectance, transmittance and absorbance

Optical Texture

Pattern observed for thin slabs of birefringent material between crossed polarizers under a microscope. It is

used as a fingerprint to help identify the many different liquid-crystal phases

Optically Detected Magnetic Resonance

Magnetic resonance observed by optical means, particularly by detection of electron spin resonance by monitoring luminescence intensity and polarization, etc.

Order Parameter

The parameter that measures the extent of long-range order characteristic of a phase. The orientational order parameter is the most important for liquid crystals; by definition it vanishes in the isotropic phase

Organic Photoconductors (OPC)

Single material or a formulated blend of materials that have photoconductive characteristics

Organic Photoreceptors

Thin-film multilayer devices made from organic photoconductive materials. These devices are often called organic photoconductors

Oscillator Strength

Measure of the probability of a transition between levels; an oscillator strength of 0.01–1 is highly probable (*allowed*) and has a short lifetime (ns), one less than 0.001 is improbable (partially *forbidden*) and has a long lifetime (μs –ms)

Oxidation-Induced Stacking Fault

A stacking fault that shows up after wafer oxidation. During oxidation, Si interstitials are injected from the wafer surface into the bulk. They aggregate around oxygen precipitates of a critical size and thereby squeeze an additional lattice plane (stacking fault) between two regular lattice planes

Oxygen Precipitates

Aggregates of oxygen atoms that form in Czochralski-grown crystals due to their relatively large oxygen content. The silica crucible that holds the Si melt is slowly dissolved during the growth process that introduces oxygen into the melt and, hence, into the growing crystal

P

Paraelectric

Nonpolar phase that transforms into the ferroelectric phase at the Curie temperature

Partial Response Maximum-Likelihood (PRML) Recording Channel

Particular form of equalization used in digital recording channels. The detector is maximum likelihood

Passivation

Passivation refers to the removal of electrical activity of a defect, often by trapping a mobile atom of the opposite electrical type (a donor may trap a mobile acceptor, for example). The defect now becomes a complex, consisting of the original defect and the

new component in close proximity, but without electrical properties. Unlike compensation, ionized impurity scattering is reduced by passivation and carrier mobility is increased

Passive Matrix Addressing

Also known as simple multiplexing, passive matrix addressing is a technique for writing images onto liquid-crystal displays. The display elements are arranged as a matrix of rows and columns, and a series of voltage pulses are applied to each row in sequence. Individual pixels are activated by applying a voltage to the relevant column, such that the sum of the row and column voltages exceeds the desired switching or threshold voltage. Each pixel responds to the root mean square voltage applied during the line-address time, and unlike the active-matrix addressed displays, there is no charge-storage facility. Thus, the number of rows that can be addressed is limited

PECVD Technique

An abbreviation for plasma-enhanced chemical vapor deposition. Like CVD, this is a technique for thin-film deposition. In contrast to the CVD technique, in the PECVD technique the deposition of a film is a nonequilibrium process. The gas-phase reactions, which are activated by temperature in the CVD technique, originate in the PECVD technique from an interaction of electrons with the gas molecules injected in the reaction chamber. In the plasma the dissociation of the gas precursors is obtained by their collisions with energetic electrons. The products of the reactions interact with the substrate surface and leads to the deposition of a thin film. As the dissociation of the gases is produced by electron impacts, the reactor and substrate can be kept at room temperature

Penetration Depth (λ_L)

The characteristic length scale for the penetration of a magnetic field into the surface layer of a superconductor

Perfect Diamagnetism (Meissner Effect)

Of the superconducting state is reflected by the fact that a magnetic field is expelled from the interior of a field-cooled superconductor as soon as the superconducting state is reached

Permalloy

Alloy of nickel and iron with approximately 80% Ni and 20% Fe. Composition of alloy with zero magnetostriction

Permittivity (ϵ)

A measure of the polarizability of a dielectric material, which is frequency dependent. The relative permittivity of a material is often given by ϵ or ϵ_r , such as with the parallel-plate capacitor expression for capacitance: $C = \epsilon\epsilon_0 A/t$. (See also *dielectric constant*.)

Perpendicular Magnetic Recording

Recording system consisting of write and read head

and a recording media where the magnetization is perpendicular to the surface of the recording medium

PES Technique (Photoemission Spectroscopy)

PES is a method to determine the barrier heights of Schottky contacts and band offsets of semiconductor heterostructures from energy-distribution curves of electrons photoemitted from Schottky contacts and semiconductor heterostructures excited by ultraviolet light (UPS), soft X-rays (SXPS) and X-rays (XPS)

Phonons

Quantized lattice vibrations in a solid

Phosphorescence

Luminescence with a lifetime $\gtrsim 10$ ns, typically involving a metastable state in the pumping cycle

Photo-Thermal Ionization Spectroscopy

A very sensitive technique for the measurement and identification of shallow impurities in semiconductors. The measurement is a two-stage process; the absorption of light excites the electron on un-ionized donors into a higher level and thermal energy now ionizes this electron into the conduction band, where it can be measured as a photo-current. Because the energy of the optical ionization is dependent on the donor type, the photocurrent spectrum as a function of photon energy, displays peaks that are characteristic of the donors present

Photo-catalysis

An enhancement of the reaction rate on the surface by shining UV radiation onto the growing film

Photochromic

Temporal coloring induced by (UV) light illumination. For instance, photo-structural changes of Ag particles, which are dispersed in oxide glasses

Photoconductivity

The part of the conductivity that is caused by the absorption of light. Negative values of the photoconductivity are possible when optical excitation activates efficient carrier traps. A photoconductive device uses the change in resistance of a slab of semiconductor to measure the extra electron-hole pairs created by the absorption of photons

Photocurrent

The current that is generated as a result of absorption of light

Photodarkening

Quasi-stable darkening induced by light illumination. In chalcogenide glasses, it occurs with a red-shift of optical absorption edge, which is induced through athermal photo-structural processes. The process, however, is speculative

Photoinduced Degradation

A deterioration of semiconductor properties by prolonged light exposure

Photoinduced Discharge Characteristic (PIDC)

Plot of exposed potential versus exposure using either

a continuous exposure or flash exposures of varying intensity

Photoinduced Effects

Changes in the properties of a glass induced by light, involving transitions between metastable states of the glass or changes in defect sites within the glass. Typically, a laser beam is used to locally modify the refractive index, density, absorption coefficient, etc. of the glass (see photodarkening, photo-induced degradation). These processes are widely used to pattern photonic structures such as Bragg gratings, waveguides, and refractive lenses into glasses

Photoisomerization

Light-induced change in the molecular structure of a molecule

Photoluminescence

Luminescence excited by the external illumination of the material. It should be distinguished from reflection and light scattering, which are also caused by external illumination and are not connected with thermal radiation of the material too. The criterion to distinguish these phenomena is the characteristic decay time after the cessation of the incident light. While reflection and scattering would decay within $\approx 10^{-14}$ s (characteristic time of electromagnetic oscillations of incident light wave) the photoluminescence would persist at least more than $\approx 10^{-12}$ s (characteristic time of atomic vibrations of luminescent material)

Photon

A quantum of electromagnetic radiation

Photonic Glass

High-purity glasses in which impurities are controlled or suppressed at ppm levels. An example is the optical fiber glass developed at the end of the 20th century, which is very recent when compared to a long history of artificial glasses of 5000 years

Photorefractive Beam Fanning

Spreading of an intense light beam due to the photorefractive properties of a material

Photosensitivity

of an organic photoconductor is a measure of the rate of photodischarge when exposed to light. Typical characterizations are the initial slope of the photodischarge (volts/J) and the energy required for photodischarge to half the initial surface potential

Photovoltaic Device

Utilizes a p–n junction in a semiconductor to separate electron–hole pairs (created by the absorption of photons) to generate a voltage

Photovoltaic Module

A complete encapsulated device suitable for mounting on a roof or building façade. Modules will often comprise of a number of cells and typical size is of the order of 1 m^2

Photovoltaic Solar Cell

A semiconductor device for conversion of solar energy into electricity

Piezoelectric Polarization

Generation of electric polarization in certain dielectric crystals as a result of the application of mechanical stress

Piezoelectricity

Property of some crystalline materials, which produce an electric charge when subjected to an externally applied force (direct effect). They also deform when subjected to an external electric field (reverse effect)

Piezoresistivity

Property of certain materials, including thick-film resistors, whereby an externally applied force gives rise to a change in resistance

Pinned Film

Coupling of a ferromagnetic to an antiferromagnetic film resulting in the magnetization of the ferromagnetic film being constrained to a fixed direction. The strength of the coupling is characterized by an exchange field parameter (H_{ua}) measured in Oe or A/m

Pixie Dust

Thin layer of Ru as used in synthetic antiferromagnetic media (SAF)

Planar Lightwave Circuit (PLC)

The industrially established processes for manufacturing integrated optics devices in silica-based glasses deposited on silicon wafers. Typically, the glass layers are deposited by chemical vapor deposition or flame hydrolysis. These technologies were developed mainly for applications in fiber optics, and are widely used to manufacture wavelength multiplexers

Plasmon

Quantized collective motions of electron gas in a metal. In a bulk metal, the plasmon propagates as a longitudinal wave, which may be probed by an electron beam. In a metal nanoparticle, a transversal surface plasmon can be excited by light waves

Plastic Deformation

Deformation of a body caused by an applied stress, which remains after the stress is removed

Plastic Electronics

Use of polymers in electronic and optoelectronic devices (See also *molecular electronics*)

Point Defect

Smallest structural element, or imperfection, to cause departure from a perfect lattice structure, e.g. a dopant or impurity atom

Poisson's Equation

A differential equation relating the spatial gradient of the local electric field to the local space-charge density

Poisson's Ratio ν

The ratio of transverse contraction strain to longitudinal extension strain in the direction of stretching force. Tensile deformation is considered positive and compressive deformation is considered negative. The definition of Poisson's ratio contains a minus sign so that normal materials have a positive ratio

Polarity

Property of a physical system that has two points with different characteristics, such as one that has opposite charges or electric potentials

Polarization Ratio

The density of states at the Fermi energy in an energy band for electrons in a metal with spins parallel (antiparallel) to the local magnetization is $D(E_F^\uparrow)$ and $D(E_F^\downarrow)$, respectively. The polarization ratio is

$$P = \frac{D(E_F^\uparrow) - D(E_F^\downarrow)}{D(E_F^\uparrow) + D(E_F^\downarrow)}$$

Poles of a Write Head

Structure in a write head that couples the magnetic fields to the write gap

Polycrystalline Solid

Polycrystalline material is not a single crystal as a whole, but composed of many small crystals randomly oriented in different directions. The small crystals in polycrystalline solids are called grains. These grains have irregular shapes and orientations. A polycrystalline material has grain boundaries where differently oriented crystals meet. Polycrystalline silicon is produced in the form of a ribbon or thin films

Polygonization

The motion of dislocations to form structures that resemble three-dimensional polygons. In this way, they minimize their strain energies. Readers are referred to the many excellent books on materials science that discuss dislocations and their interactions

Polymer Thick Films

are those that are cured at much lower temperatures than their cermet counterparts. The binding matrix is a polymer material and the resulting films can be used in flexible circuits

Positron Annihilation

Analytical technique that is sensitive to open defects, such as vacancies, in a crystal. The lifetime of positrons is greater when the total number of electrons is reduced, as happens at vacancies. Measuring positron lifetime can give semiquantitative estimates of the concentrations of these defects

Power Stabilizer

Optical device that can provide a fixed output intensity upon varying input intensity. The principle may be based on multi-photon absorption, which becomes efficient when light is more intense

Precursor

Volatile compound containing the element required for deposition onto the substrate

Process Integration

The carefully engineered combination of materials and processes to enable the fabrication of an integrated circuit technology. The careful consideration of materials properties, device physics and electrical engineering principles is required for successful integration of materials into an IC chip

Proper Ferroelectric

Ferroelectric material in which the spontaneous polarization is the primary order parameter

Pseudo-morphic Layers

True epitaxy only occurs when the lattice constants of the substrate (material A) and layer (material B) are equal. If they are different, misfit dislocations or even gross bending of the structure can occur. In pseudo-morphic growth, the lattice constants can be very different. However, before misfit dislocations can be formed, either the growth of B is stopped or a layer of the substrate material, A, is grown on top of the structure. In either case, the structure is stabilized with material B being severely strained. Manufacturers use this technique to produce devices with advanced electrical or optoelectronic properties

p-Type Conductivity

Semiconductor material in which the dopants create holes as the majority charge carrier. It is formed by doping with acceptor atoms

Pulse Width

Width of a pulse in a digital magnetic recording read head when sensing a recorded transition. Characterized by the 50% width (PW_{50})

Pyroelectric Effect

Generation of an electric polarization change or a charge separation in a material in response to a change in material temperature

Pyrolytic Boron Nitride (pBN)

This is the favored crucible material for GaAs growth from the melt and for Knudsen cells for MBE. Pyrolytic refers to its high-temperature capabilities

Q**Quantum Cascade (QC) Laser**

Light amplification by stimulated emission of radiation (laser) through inter-subband transitions and unipolar electron transport

Quantum Confined Stark Effect

The change in near-band-edge optical absorption in quantum wells that takes place when a high electric field is present

Quantum Dot (QD)

A semiconductor quantum heterostructure that is quantum confined in all three dimensions, for example

InAs islands embedded in a layer of GaAs and Ge islands in Si

Quantum Efficiency

Indicates either the probability that the absorption of a photon will result in the creation of a free electron and hole or the ratio of the number of luminescence photons to the number of stimulating photons. When determined by xerographic discharge of the photoreceptor QE is a measure of the overall effectiveness of surface charge neutralization per absorbed (or incident) photon. It is sometimes called xerographic gain or supply efficiency. When determined by spectroscopic methods, such as fluorescence quenching, it is a measure of charge generation within the light-absorbing moiety

Quantum Well (QW)

Low-dimensional semiconductor quantum heterostructure where a thin layer with lower bandgap is sandwiched between layers (barriers) with higher bandgap, leading to spatial confinement of carriers along one dimension for which the energy is quantized in discrete levels. Carriers have free motion in the plane perpendicular to the confinement direction. May be repeated to produce a multiple quantum well (MQW)

Quantum Well Inter-subband Photodetector (QWIP)

Long-wavelength (IR) detector based on light absorption through inter-subband electron transitions in n-doped QWs

Quantum Wire

Semiconductor quantum heterostructure with two-dimensional confinement of carrier motion.

Carriers have just one direction of free motion

Quantum-Confined Stark Effect (QCSE)

Response of the confined electrons and holes in a quantum well to a strong DC electric field applied in the growth direction

Quantum Size Effect

Quantum effect associated with quantization of the band and level. It is observed in two-, one- and zero-dimensional systems whose size becomes small

Quasi Particles

Unpaired electrons excited above the energy gap

Quasi-Fermi Levels

Levels that correspond to the energy positions the equilibrium Fermi level would need to have in the gap in order to produce equivalent carrier densities to the ones that are generated by the illumination

R

Radiation Resistance

This is important for solar cells operating in space where high doses of gamma rays and high energy particles can significantly degrade the efficiency of the cells

Radio-Frequency (RF) Circuits

High-speed and microwave analog circuits that operate in the range 0.1–100 GHz

Reactive Sputtering

Technique for preparing thin films by sputtering with gas mixture of reactive gas and conventional sputtering gas such as oxygen and argon

Reactor Cell

Chamber where the precursors react to deposit a film onto the substrate

Read Head

In magnetic recording the ferromagnetic device used to generate a voltage proportional to the state of magnetization in the recording medium

Reciprocal Lattice

Theoretical lattice constructed from a real lattice, such that any vector from the origin to a diffracted spot is normal to a particular plane in the real lattice, with reciprocal length of that plane spacing

Recombination

Process whereby non-equilibrium populations of electrons and holes return to their equilibrium values. Depending on whether the recombination rate is proportional to the excess carrier density or the square of that quantity, the processes can be referred to as linear and quadratic, or alternatively as monomolecular and bimolecular recombination. Radiative recombination refers to the transfer of at least some of the energy of the excess carriers into photons. Non-radiative recombination involves only the production by lattice phonons (heat)

Reflectometer

Instrument used for reflecting light off the substrate to monitor the growth process. This is normally achieved through detecting the interference modulation of the light intensity from a growing film

Reflow

Process of heating a glass above its glass-transition temperature, to the point that its viscosity is sufficiently reduced to enable the material to flow. In combination with surface tension effects or other external forces, reflow is often exploited in the reshaping of optical devices

Remanent Magnetization

is magnetization that remains in a sample after the magnetic fields are removed

Remanent Polarization

Dielectric polarization that remains in a ferroelectric material after an electric field has been applied

Resistivity

Parameter of a semiconductor that depends on the free electron and hole densities (cm^{-3}) and their respective mobilities (cm^2/Vs) and is expressed in units of $\Omega \text{ cm}$. It is the reciprocal of conductivity and it depends strongly on temperature in a semiconductor

Resolution

The smallest separation of two points in an object that can be distinctly reproduced within an image

Resonant Nonlinear Response

The nonlinear response taking place in the absorbing spectral region

Responsivity

Signal term often measured using a two-temperature black body. Units can be V/W (used for photoconductors) or V/photon (often used for multiplexed photodiodes)

Retrograde Solidus

Describes the shape of the solidus when it shrinks as the temperature is reduced. In GaAs, the increased width of the solidus at high temperatures indicates that concentrations of Ga or As, in excess of stoichiometry, can exist in the solid crystal. These concentrations must reduce as the crystal cools because of the retrograde solidus

RHEED

Glancing-angle electron diffraction technique, sensitive to surface reconstruction and morphology. Key in situ analytical technique in MBE

ROIC

Abbreviation for readout integrated circuit, commonly used to describe the silicon chip on which the detector material is mounted. The role of the ROIC is to integrate the signal, perform some signal processing and readout the array. Other terms are: multiplexer or mux

R-T Method

Defining the optical constants, based on two equations connecting the formers to the reflectance and transmittance of the slab sample of material under measurement

Rutherford Scattering

Elastic scattering of electrons due to an electrostatic interaction with the nucleus and surrounding electrons of an atom

S

Saturated Vapor Pressure (SVP)

Partial pressure of a substance in equilibrium with its liquid, so the partial pressure is determined solely by the temperature of the liquid and vapor

Saturation Intensity

Intensity at which the effective absorption decreases to a half of its initial value

Saturation Magnetization

Maximum magnetization that a ferromagnetic sample can attain in magnetic fields larger than the coercive field

Saturation of Absorption

Resonant nonlinear process in which absorption decreases with increased level of illumination

Scaling

Calculated reduction of integrated circuit element dimensions according to physical and engineering principles and constraints, as well as economic considerations (See *Moore's Law*)

Scanning Probe Microscopy

Generic term given to microscopy techniques that use a scanned micro- or nanoscale tip in immediate proximity to a surface to image topography or other physical features with almost atomic resolution. The main types are atomic force microscopy (AFM) and scanning tunneling microscopy (STM)

SCH Laser

Separate confinement heterojunction laser. In the SCH laser the optical and electrical confinements are achieved separately by altering the thickness and the alloy composition of the cladding layers

Schottky Contacts

Metal-semiconductor contacts are also named Schottky contacts

Schottky-Mott Rule

Schottky-Mott rule equates the barrier heights of n-type (p-type) Schottky contacts with the difference of the metal work function and the electron affinity (ionization energy) of the semiconductor. The rule is incorrect since it does not consider the *intrinsic* interface-induced gap states

Screen

Mask used to define the desired thick-film pattern. It is usually made of stainless steel, polyester or nylon

Screen Printing

Method by which thick films are deposited onto substrates. See also *antiferromagnetically coupled media* (AFM)

Secondary Electrons

Low-energy (< 50 eV) electrons that escape from the near sample surface, used to form topographic images

Secondary-Ion Mass Spectrometry (SIMS) and Glow Discharge Mass Spectrometry (GDMS)

Mass-spectrometric techniques that are well suited for the chemical analysis of semiconductors. In addition to being extremely sensitive to most impurities, they are quantitative and can give valuable information regarding the distributions of concentrations with depth. In SIMS, high-energy primary ions are focused on the sample surface. These sputter atoms from the material under investigation. Those that are charged are passed into a mass spectrometer for analysis. In GDMS, the sputtering is accomplished by a glow discharge. SIMS is, by convention, generally classified as dynamic, in that the material surface layers are continually removed as they are being measured, and static, in which the ion dose during measurement is restricted to less than 10^{16} ions/m² in order to retain the surface in an essentially undamaged state

Segregation Coefficient, k

In growth from the melt the incorporation of an impurity into the crystal depends on the equilibrium ratio of the solubilities in the melt and the solid. For a dilute solution, this ratio is given by the ratio of the equilibrium value of impurity concentration at the solidus and the liquidus at the growth temperature. This ratio is the segregation coefficient, k . For most impurities, k is less than unity and the growing crystal contains a lower impurity concentration than in the source melt. The exploitation of this fact has resulted in the purification technique of *zone refining*

Selection Rule

Quantum-mechanical rule based on wave-function and operator symmetry that determines oscillator strength; typically set by parity or angular-momentum considerations

Selective Epitaxy

Growth of a single-crystal layer in a window without any deposition on the surrounding mask layer

Selenization

Process of forming the CIGS or CIS layer by exposing a copper, indium and gallium precursor layer to Se to form the alloy

Self-Assembly

Spontaneous formation of a layer of organic material on a solid substrate surface. The process usually takes place by immersing the substrate into a solution of the organic compound. The driving force for the self-assembly can be chemical and/or electrostatic attraction

Self-Defocusing

A nonlinear process associated with negative refractive nonlinearity that results in spatial spreading of intense light

Self-Focusing

A nonlinear process associated with positive refractive nonlinearity that results in focusing of intense light

Semi-Insulating

This is a term given to semiconductors whose resistivity lies above about $10^6 \Omega \text{ cm}$. In GaAs, semi-insulating properties normally lie in the range of $10^7 - 10^8 \Omega \text{ cm}$

Semiconductor Detector

Device that converts the incident photons directly into electrical pulse

Semiconductor Nanocrystals

Composite materials made out of semiconductor quantum dots embedded in organic or glass hosts

Shallow-Energy-Level Dopant

Doping impurities whose energy level lies very close to the conduction or valence band for donors or acceptors

Shear Modulus

Sometimes also called the rigidity, relates stress and strain according to Hooke's law and is a measure of

a material's resistance to shearing stress. The shear modulus therefore has units of pressure

Sheet Resistance

Resistivity of a semiconductor sample divided by its thickness, measured in Ω / square . It is commonly used by integrated circuit designers when designing resistances by specifying the number of required squares required to give a certain value of resistance

Shields

Soft ferromagnetic films used to direct the flux from a recording layer away from sensor films in a read head

Short-Range Atomic Structure

Atomic bonding structures within a scale of $\approx 0.5 \text{ nm}$, which are characterized by coordination number (the number of nearest-neighbor atoms), bond length, and bond angle. It is demonstrated for such simple glasses as SiO_2 that the short-range structure is nearly the same with that in a corresponding crystal

Signal Decay Rate

Rate (R) at which the amplitude of a signal read back from a recording system decays with time

$$R = 100 \frac{\frac{A(t)}{A(t_0)} - 1}{\log(t/t_0)}$$

Single-Layer Organic Photoconductor

A photoconductor with an architecture where the charge-generation and charge-transport functions are combined into a single layer

Slider

In a disk drive, the structure used to support on an air bearing the write and read heads over the recording medium

Sliding Boat

Liquid-phase epitaxy technique in which the substrate is slid under the melt in a horizontal orientation

Slope Parameter

(In perpendicular magnetic recording) The slope of the magnetization curve at the coercive field times

$$4\pi\alpha = 4\pi \frac{dM}{dH(H=-H_c)}$$

Slush

Homogeneous charge of a ternary is held across the liquidus–solidus gap with the lower end solid, the upper end liquid and the central section in a slushy state, during recrystallization

Smectic Phase

Liquid-crystal phase with some long-range translational order in addition to the long-range orientational order of a nematic. The constituent rod-like molecules are arranged in layers giving translational order in one dimension

Soft Underlayer (SUL)

Magnetically soft (low-coercivity) film underneath the recording layer in a perpendicular magnetic recording system. Used as a low-reluctance path for the flux from the write head

Solar Cell

Semiconductor device that converts the energy of sunlight into electric energy. Also called photovoltaic cell

Solid/Liquid/Gaseous Phase Equilibria

These phase equilibria are essential for understanding both the various growth techniques used and in understanding post-growth heat treatments, i.e. cool-down after growth and subsequent annealing stages

Solidus

In an equilibrium phase diagram, the solidus is the line below which all the components are solid

Solitonic Propagation Of Pulses

Propagation of pulses characterized by a lack of temporal pulse spreading

Spatial Frequency

Spatial frequency is the reciprocal of a spatial dimension (e.g., height x or width y in two-dimensional image) similar to the temporal frequency, which is the reciprocal of time. However, the unit of spatial frequency is preferably expressed as line pairs/mm (lp/mm), instead of cycles/mm

Specific Heat

(also called the specific heat capacity). Amount of heat required to change a unit mass (or unit quantity, such as mole) of a substance by one degree in temperature

Spherical Aberration

A blurring of resolution due to the spread of path lengths of rays traveling from an object to the image plane, arising from a variation of the focal length of a lens as a function of distance from the center of the lens

Spin Valve

Read head that uses the giant-magnetoresistive effect in three metallic films: the ferromagnetic free film, a non-magnetic spacer film and a ferromagnetic pinned film

Spontaneous Polarization

The electric polarization that a substance possesses in the absence of an external electric field

SPRITE

Abbreviation for signal processing in the element: a device used in the UK common module camera. It relies on a strip of CMT with a high voltage bias to drift photon generated holes at the same speed as the image is scanned, so resulting in an amplified signal at the end of the strip

Sputter Depth Profile

Compositional depth profile obtained when the surface composition is measured as material is removed by sputtering. Note: in some analytical methods such as SIMS, the sputtering is often accomplished by the ion beam used for analysis, but in other methods an ion beam may need to be added

Sputtering

Process in which atoms and ions are ejected from the sample as a result of particle bombardment

Squeegee

Device used to transfer the thick-film paste through a screen and onto the substrate

SRAM

Static random-access memory

SSR

Solid-state recrystallization is a growth technique that produces a homogeneous but polycrystalline charge that is subsequently recrystallized in the solid state to produce multi-grained material

Stepanov Technique

Crystals are pulled from a crucible containing a crystal-shaped aperture

Stoichiometry

State of a perfect compound where the ratio of the numbers of atoms of the elements is a simple fraction. In GaAs, for example, stoichiometry exists when the number of As and Ga atoms are equal

Stokes Shift

Difference in energy between the maximum of the emission spectrum and the maximum of the excitation spectrum

Strain (ϵ)

Defined in elementary form as the change in length divided by the original length., it can be thought of as the movement of one corner of a cubic box from its initial position under stress

Strained Quantum Well

QW layer grown on a substrate with a different lattice constant, resulting in a significant strain of the QW layer lattice

Stress

Force per unit area provided either by gravity or by the flow of viscous fluid

Strong Anchoring

Strong anchoring of the director at a surface parallel to the easy axis is said to occur when an applied field is unable to alter the orientation of the director at the surface but can in the bulk

Structural Relaxation

Essentially an aging effect associated with glasses. Because glasses are metastable materials with random network structures, they are inherently subject to short- or long-term changes in material properties. Often, structural relaxation is manifested by a change in specific volume (densification) at fixed temperature versus time. The rate of such changes is extremely sensitive to the difference between the glass transition temperature and the observation temperature. Structural relaxation can be induced rapidly by an annealing step, in which the glass is heated near its glass-transition temperature for some period of time

Substitutional Impurities

Impurities that replace the crystal's base atom at that base atom's lattice position

Substrate

Base material onto which a film is deposited.

Examples of typical substrates include materials such as alumina, beryllia, aluminum nitride, silicon, insulated steels and various plastics. In epitaxial processes, substrates are normally near lattice-matched pieces of material of a sufficiently large area for use in the various epitaxial growth processes, normally from similar, i. e. common-cation, ternary systems, e.g. CdZnTe for CMT

Superconductors of Type I

Characterized by the fact that normal and superconducting regions cannot coexist in these materials. The value of $\kappa = \lambda_L/\xi$ is smaller than $1/\sqrt{2}$

Superconductors of Type II

Type II superconductors, for which $\kappa = \lambda_L/\xi > 1/\sqrt{2}$, show perfect diamagnetism only below the lower critical field B_{c1} . Between the lower and the upper critical fields type II superconductors are in the mixed state. Magnetic flux enters the superconductor via the normal conducting cores of the flux lines

Supercooling Temperature

Difference between the glass-transition temperature and the in-use temperature for a glass-based device. For a large (small) supercooling temperature, the structural relaxation rate is low (high)

Superlattice (SL)

Structure of repeated QWs with thin barriers allowing the coupling of wave-functions from adjacent QWs and the subsequent delocalization of the energy levels

Superlattice Avalanche Photodiode (SL-APD)

Avalanche photodiode architecture where an SL is incorporated in the carrier multiplication region for the purpose of reducing the dark current as well as the excess noise factor

Superparamagnetism

In small ferromagnetic particles the magnetization can flip from one state to the opposite state under thermal excitation. This behavior is similar to a paramagnet. The resistance of the particle to switching is characterized by a thermal stability factor $= \frac{K_u V}{k_B T}$, where K_u is the uniaxial anisotropy factor for the particle, V is the particle volume, k_B is the Boltzmann constant, and T is the absolute temperature

Supersaturation

Ratio between the species concentration and its solubility limit/equilibrium concentration at a given temperature when this ratio is larger than 1

Superstrate

Describes the thin-film configuration where the glass substrate acts as the window for solar radiation and therefore needs a TCO layer before the photovoltaic structure can be deposited

Surface-Mount Devices (SMDs)

Electronic components that are attached to the surface of a circuit board as opposed to having through-hole connections. They are a characteristic feature of a hybrid circuit

Surface Passivation

Semiconductor surfaces are often electrically active and appear to be covered with a high density of deep-level states. These can greatly affect the properties of a device. Fortunately, it is often possible to treat the surface to substantially reduce this density to values that do not affect device operation. This is surface passivation. In GaAs, passivation is often accomplished by covering the surface with a layer of GaAlAs. Also, see passivation

Susceptor

This is normally made of high-density graphite and is used to transfer the heat from the heater (possibly RF coupling straight into the susceptor) to the substrate

SW

Abbreviation for short wave, often used for wavelengths between $1.0\ \mu\text{m}$ and $3\ \mu\text{m}$. The atmosphere is transparent in relatively narrow bands within the SW region, the most common one is between 2.0 and $2.25\ \mu\text{m}$, although $1.5\ \mu\text{m}$ is also important as this is the wavelength for eye-safe lasers

Synthetic Antiferromagnet

Sequence of films: antiferromagnet/ferromagnet/ruthenium/ferromagnet. An example is: MnFe/Co/Ru/Co. With a thin Ru film, the final ferromagnetic film is antiferromagnetically coupled to the first ferromagnetic film. The coupling between the two ferromagnetic films is characterized by the exchange coupling parameter J_{12} (erg/cm^3 or J/m^3)

Synthetic Ferrimagnetic Media (SFM)

Magnetic recording media in which there are two ferromagnetic layers of unequal thickness coupled by a thin layer of ruthenium (Ru), sometimes referred to as "Pixie Dust". The two layers are antiferromagnetically coupled. The coupling between the ferromagnetic films is characterized in terms of the antiferromagnetic coupling between the thinner layer (layer 2) and Ru and the media flux from that layer by an exchange magnetic field $H_{\text{ex}} = J_{\text{ex},2}/M_2 t_2$

T**T Nonlinear Figure of Merit**

Figure of merit that describes the applicability of a refractive nonlinear optical material in terms of the Kerr coefficient and two-photon absorption

Tail States

Localized states in the tail of band, i. e. conduction band and valence band, which generally have an exponential density-of-states function

Tauc Gap

Tauc discovered that, in many chalcogenide glasses such as As_2S_3 , optical absorption spectra α around the fundamental edge can be fitted as $\alpha\hbar\omega \propto (\hbar\omega - E_g)^2$, where E_g is called the Tauc gap. The energy is often used as a measure of optical bandgap, while its theoretical interpretation is not conclusive

Technology Node

Minimum half-pitch of metal interconnect is most representative of the process capability enabling high-density (low cost/function) integrated circuits and is selected to define an ITRS technology node. For each node, this defining metal half-pitch is taken from whatever product has the minimum value. Historically, DRAMs have had leadership on metal pitch, but this could potentially shift to another product in the future. Other parameters are also important for characterizing IC technology. For example, in the case of microprocessors (MPUs), physical bottom gate length is most representative of the leading-edge technology level required for maximum performance. Each technology node step represents the creation of significant technology progress in metal half-pitch – approximately 70% of the preceding node, 50% of two preceding nodes. (From [IRTS 2003])

Temperature Coefficient of Resistance (TCR)

Denotes the sensitivity of a resistor material to changes in temperature. It is usual to quote TCR in terms of parts per million (ppm) per $^{\circ}\text{C}$

Tensile Strain

Type of strain obtained when a strained Si layer is grown on a relaxed $\text{Si}_{1-x}\text{Ge}_x$ layer

Ternary and Quaternary Alloys

These are alloys containing three or four components, respectively. GaAlAs and GaAlAsP are examples

Thermal Budget

Term describing the temperature–time product associated with an IC annealing process step. Material stability and morphology is typically very sensitive to annealing temperature for a period of time

Thermal Conductivity

(heat conductivity) Heat flow across a surface per unit area per unit time, divided by the negative of the rate of change of temperature with distance in a direction perpendicular to the surface

Thermal Expansion Coefficient

Fractional change in length or volume of a material for a unit change in temperature

Thermal Nonlinearities

Nonlinear effect associated with heating of the materials by intense light

Thermally Stimulated Current (TSC) spectroscopy

TSC is a useful technique for assessing concentrations of deep levels in high-resistivity semiconductors. These levels are filled optically with light above the bandgap energy. They are emptied sequentially, with

the sample under bias, as the sample temperature is raised. Defects of different ionization energies produce separate peaks in the graph of current as a function of temperature

Thermistors

Thermally sensitive resistors that exhibit a change in resistance when the temperature is altered. The most common forms have a negative TCR, meaning that the resistance decreases as the temperature increases

Thermomechanical Fatigue

Occurs when materials with different CTE are joined and used in an environment that experiences cyclic temperature fluctuations resulting in imposed cyclic strain that results in damage to the joined materials

Thermophotovoltaic (TPV)

Absorption of solar radiation and re-emission of infrared radiation that is then converted into electricity by absorption in a narrow-bandgap cell

Thermotropic

Liquid crystals are those where the transition from one phase to another occurs on changing the temperature

Thick Film

Layer deposited onto a substrate by the process of screen printing

Threading Dislocations

This term is applied to dislocations that are formed at an interface and that *thread* their way into the epitaxial layer

Threshold Voltage (V_{th})

Voltage at which an inversion layer forms in the semiconductor substrate of an MIS structure. For applied voltages beyond this threshold, the transistor turns on, i. e. a conducting channel is established between the transistor source and drain, as observed by the increase in drive current

Time Division Multiplexing

Multiplexing technique used in modern optical networks that allows close spacing in time of bits in a single channel

Time-Domain Charge Measurement (TDCM)

TDCM is a rapid technique for the non-contacting measurement of electrical resistivity in SI materials. Its speed and high spatial resolution allows TDCM to be used in a mapping system

Tipping

Liquid-phase epitaxy technique in which the melt is moved over the substrate in a tipping furnace

Transferred Electron Effect

This is a somewhat unusual effect where conduction-band electrons can exist in alternative conduction-band states. In GaAs, the lowest conduction-band minimum corresponds to zero electron momentum. The application of an electric field can excite electrons into subsidiary minima of somewhat greater energy from where they will relax back to their original states. If the effective mass of the

electrons in the subsidiary minima is greater than that in the primary minimum, negative differential resistance can be realized. This effect can be found in GaAs, InP and certain alloys

Transistor

A three-terminal device in which the current flow between two terminals (called the source and drain regions) is controlled by the voltage applied between a third terminal (the gate electrode) and one of the two terminals

Transit Time

Refers to the difference in time between the moment a charge carrier is generated at one end of the sample and its arrival at the other end. It is the primary result in time-of-flight experiments. In xerography, transit time is the time for a photoinjected carrier to traverse the charge-transfer layer

Transparent Conducting Oxide (TCO)

Thin layer of highly conducting material that is used as the front contact in thin-film solar cells. The requirement of high optical transmission over a wide spectral range is important to allow as much of the sunlight through to the absorber layer

Transverse Magnetic Bias Film

In a magnetoresistive head, an antiferromagnetic film coupled to the pinned film to maintain the magnetization of the pinned film in a direction transverse to the magnetization of the recording film

Trapping Centers

Irregular sites in the photoconductor with localized electronic energy levels inside the bandgap. Such sites will trap charge carriers and thus inhibit electrical conduction

Traveling Heater Method

A molten zone is made to migrate through a homogeneous solid source material

Trimming

Process by which the value of thick-film elements can be adjusted. Usually achieved by using a laser or air abrasive jet

Tunneling Magnetoresistance (TMR)

With two ferromagnetic films coupled by a thin insulating layer, electrons can tunnel through the insulating layer and the magnetoresistance coefficient is given in terms of the polarization ratio for the two contacts as $\frac{\Delta R}{R} = \frac{2P_1 P_2}{1 - P_1 P_2}$.

Tunneling Conduction

Process for charge conduction where the charge carriers pass through an energy barrier by quantum-mechanical tunneling

Twin Crystal

Crystal having two or more crystals or crystal sections that, when regularly positioned, are in reverse position relative to the other sections

Two-Photon Absorption

Nonlinear process in which a simultaneous absorption of two photons results in an electronic transition

U

UHV

Ultra-high vacuum is a vacuum better than 10^{-9} Torr

Ultrafast Nonlinear Response

Nonlinear response taking place in the non-absorbing spectral region

Underfill

Dielectric composite organic material that is bonded between the chip and substrate of a flip-chip device to help mechanically interlock the chip to the substrate. The underfill material is typically a silica-filled anhydride resin polymer

Uniaxial Anisotropy

Magnetic anisotropy along one direction in a ferromagnetic material and characterized by a uniaxial anisotropy parameter K_u (erg/cm³ or J/m³)

Uniform Planar Alignment

of the director is when the director is parallel to the surface and to a particular direction in the surface. This has also been referred to as homogeneous alignment

Unipolar Avalanche Photodiode (UAPD)

QWIP detector employing avalanche multiplication of only one type of carrier via intra-QW impact ionization by carrier-carrier scattering

Upper Critical Field (B_{c2})

The highest magnetic field allowing the existence of the superconducting mixed state in the type II superconductor considered

V

V/I boundary

Denotes the spatial location of the transition from the vacancy-dominated region to the Si interstitial dominated region and vice versa

Vacancy

Regular lattice site from which the host atom is missing. Anion (cation) vacancies are generally donors (acceptors)

Valence Band

Highest range or band of energies in a semiconductor where electrons are normally present at zero temperature. When electrons are promoted from this band, holes are left behind that contribute to the electrical conductivity

Vapor Growth Techniques

Growth takes place via a vapor phase: particularly useful for high-melting-point materials, or those with high partial pressures of one constituent, e.g. ZnS

Vapor-Phase Epitaxy (VPE)

This is a form of epitaxy where the components of the layer to be grown are transported to the substrate as a vapor. Decomposition of these components to produce the layer occurs because the substrate is heated, often on a support called a susceptor. The different types of VPE are discussed in the text

(VCSEL)

Vertical-cavity surface-emitting laser. One of the more modern forms of semiconductor laser, employing the high gain of quantum wells in a low-loss optical cavity

Vegard's Law

States that the lattice parameter of an alloy material is given by the concentration weighted average of the constituents

Vehicle

Liquid component of the thick-film paste. Typically contains a resin dissolved in a solvent. It is removed during the drying and firing processes

Verneuil Technique

Rapid growth method used for many high-melting-point materials, mainly oxides

Vertical-Gradient Freeze Technique

Similar to Bridgman but freezing is controlled by moving a temperature gradient along a stationary crucible

Vertical Transport

Mechanisms of carrier transport parallel to the growth direction in a semiconductor quantum-confined structure

Vertical-Cavity Surface-Emitting Laser (VCSEL)

QW diode laser emitting through its top semiconductor surface

VI Recombination

Mutual annihilation between vacancies and Si interstitials

Virtual Gap States

The virtual gap states are the *solutions* of Schrödinger's equation for complex wave vectors in the energy gap. The continuum of these *virtual* states is the source of *real* impurity states in the bulk as well as of surface states and the interface-induced gap states at *real* surfaces and interfaces, respectively, provided the corresponding boundary conditions are considered

Void

see COP

VPE

A general term used to describe the deposition of an epitaxial thin film from the vapor phase

W**W Nonlinear Figure of Merit**

Figure of merit that describes the applicability of a refractive nonlinear optical material in terms of nonlinear index change and linear absorption

Wavelength Division Multiplexing

Multiplexing technique used in modern optical networks that involves sending many signals in parallel at closely spaced wavelengths along the same fiber

Weak Links

Large-angle grain boundaries in cuprate superconductors, which act as barriers for the supercurrents

Web Photoreceptors

Devices where the layers are coated on an insulating polymeric support and subsequently fashioned into a loop

Wire Bonding

Process where a thin wire (can be less than 25 μm) made of Au or Al is bonded to the surface of an integrated circuit and then to a pad or a leadframe in the package

Work Function

The energy (usually measured in electron volts) needed to remove an electron from the Fermi level in a solid to outside the surface

Write Efficiency

Ratio of the magnetic field times write gap to the magnetomotive force (turns times current) in the write coil

Write Gap

Region in a write head that generates the magnetic fields that couple to the recording medium. Characterized by the gap length g

Write Head

In magnetic recording the ferromagnetic device used to generate magnetic fields from current that can switch the state of the magnetization in the recording medium. In disk drives the ferromagnetic device is made using thin films

Wrong Bond

Homo-polar bonds in stoichiometric glasses, such as As–As in As_2S_3 . The defective bond nominally does not exist in the corresponding crystal. However, specifically in covalent glasses such as As_2S_3 , the bond exists with a concentration of $\approx 1\%$, which depends upon preparation methods and so forth

Würtzite Structure

Comprises two interpenetrating close-packed hexagonal lattices, one for cations and the other for anions. Each anion (cation) has four cation (anion) nearest neighbors. In principle diffusion should be anisotropic but meagre available evidence indicates only slight effects

X**Xerographic Discharge**

Time-dependent decrease in surface potential with exposure of a charged OPC

Xerographic Gain

or quantum efficiency of supply is the fractional number of surface charges neutralized per absorbed photon

Xerography

Name coined by Chester Carlson for electrophotography using dry powder marking particles

X-Ray

Form of energetic electromagnetic radiation of wavelength ≈ 0.1 nm

X-Ray Diffraction

A highly sensitive technique for measuring the lattice constant of crystalline solids, see B K Tanner and D. K Bowen, 1980

X-Ray Photoelectron Spectroscopy (XPS)

Method in which an electron spectrometer is used to measure the energy distribution of photoelectrons and Auger electrons emitted from a surface irradiated by X-ray photons

X-Ray Sensitivity

The X-ray sensitivity of a photoconductive detector is the collected charge per unit area per unit exposure of radiation

X-Ray Topography

This is a method for sampling the diffraction condition of X-rays from a surface in order to observe changes in lattice constant. The X-rays are scanned over the surface and an image is built up from changes in the

fraction diffracted at a particular angle. Changes in lattice constant resulting from strain or changes in composition can be imaged in this way

Y**Yield Strength**

Applied stress (in pounds per square inch or psi in the English system, megapascals or MPa in the metric system) under which an object experiences plastic deformation

Young's Modulus

Ratio of a simple tension stress applied to a material to the resulting strain parallel to the tension. Therefore Young's modulus has units of pressure

Z**Zinc-Blende Structure**

A crystal structure that has two interpenetrating face-centered cubic lattices, one for the cations and the other for anions. Each anion (cation) has four anion (cation) nearest neighbors. Diffusion is isotropic

Zone Refining

Technique used to repeatedly pass zones of molten material through a solid bar in order to purify it, either for use directly in applications or to produce pure starting materials, e.g. elements for compound semiconductors

Subject Index

π band 1333
 π bond 1333
 π -electron delocalisation 1256
 π^* band 1333
 σ bond 1333
 μ c-Si:H
 – dangling-bond-defect 581
 – dangling-bond-defect density 589, 590
 – device applications 590
 – device-grade 582
 – epitaxial-like crystal growth 588
 – film 588
 – formation 585
 – growth 582, 585, 587
 – growth process 581
 1-D photonic crystal 999
 2-D electron gas (2DEG) 39
 2DEG heterostructure 40

A

Abbe number 74, 75, 1066, 1333
 about disordered
 – ionic conduction 161
 abrasive trimming 725
 absorption 1064
 – linear 1067
 – rate 891
 – saturation 1079
 absorption coefficient 190, 1086, 1333
 – amorphous semiconductors 570
 – effective 1064
 AC Josephson effect 1200
 accelerated crucible rotation technique (ACRT) 236, 305
 accelerometers 729
 acceptor 1333
 – dopants 125
 – doped oxide defect diagram 222
 – impurities 216
 acceptor concentration 1333
 acceptor level 1333
 acceptor-doped material 222
 acceptors, group V 850
 ACCUFET (accumulation-mode MOSFET) 550
 accumulation 1333
 acoustic-phonon scattering 465, 767

ACRT 1333
 activation energy 166, 217, 402, 845, 1113
 active components 725
 active material 1333
 active matrix addressing 937, 1333
 active matrix array (AMA) 1121
 active matrix flat-panel imagers (AMFPI) 1121
 active region 914
 actuator 728, 1152
 AC conductivity 189
 adhesive
 – interconnects, rework 1283
 – isotropic 1282
 a -face growth 543
 aggregation temperature 110, 114
 Ag-In-Sb-Te (AIST) 1140
 Aharonov–Bohm effect 1150
 air mass (AM) 1096
 air mass number 1333
 AlGaAs 8, 10, 11
 AlGaAs alloys 510
 AlGaInp 11
 AlGaN 8, 12, 13
 – alloy 792
 AlInGaAs, AlInGaP 8
 AlInN/AlGaN lattice-matched 793
 alkali growth 235
 all-optical switch 1333
 alloy
 – binary, solid solution 26
 – composition
 – interpolation scheme 736
 – disorder scattering 40
 alloy semiconductor
 – optical spectra 748
 alloys 1155
 – Group III–V 739
 – Group I–VII 739
 AlN (aluminium nitride) 805, 820
 – electrical properties 773
 – electron transport 817
 – film 760, 773, 788
 – fundamental properties 755
 – heteroepitaxial films 777
 – lattice constants 757
 – material parameters 809
 – mechanical properties 759
 – nucleation 769

– optical phonon frequencies 760
 – optical properties 786
 – single crystals 773
 – thick film technology 720
 alumina
 – materials 720
 – self-ordered porous 1005
 – tape growth 236
 – thick film substrates 720
 aluminium oxide
 – DRAM capacitor 649
 ammonia sensor 1243
 AMOLED display 1107
 amorphous
 – As₂Se₃ 140
 – organic semiconductors 1113
 – polymer 394
 – selenium 146
 – semiconducting film 1107
 amorphous film 683, 687
 – electronic properties 1108
 – metallic 688
 amorphous metaloxides 1333
 amorphous network 1334
 amorphous semiconductor 59, 99, 144, 565, 956, 1067, 1114
 – absorption coefficient 570
 – conduction band 566
 – electrical conductivity 573
 – electrical Properties 573
 – electronic states 566
 – Hall effect 574
 – light-induced phenomena 575
 – linear 1067
 – optical properties 570
 – photoluminescence 570
 – photoluminescence spectra 65
 – refractive index 1067
 – stationary photoconductivity 179
 – Structural properties 568
 – thermoelectric power 574
 – valence band 566
 amorphous semiconductors 1334
 amorphous silicon (a-Si:H) 140, 566, 587, 1334
 – dangling-bond-defect 581
 – dangling-bond-defect density 589
 – device applications 590
 – device-grade 582

- drift mobility 574
- electrical conductivity 574
- electronic material properties 1107
- electronic transport properties 1107
- growth 582, 585
- growth process 581
- hydrogenated 581
- optical properties 571
- sputtered 1114
- structural models 566
- TFT 1107
- amorphous tantalum pentoxide
 - capacitor 649
- amorphous-to-crystalline transition 687
- amperometric sensor 224
 - charge transfer 1250
 - sensitivity 1250
- amphiphilic 1334
- amphoteric dopants 1334
- amphotericity 1334
- amplifier 9
 - optical 888
- amplitude reflectance 695
- analytical electron microscope (AEM) 343
- analytical electron microscopy 1334
- Anderson model 165
- Anderson's criteria 1109
- angular dispersive X-ray diffraction (ADX) 758
- anion 1334
- anisotropic conductivity 1284
- anisotropy
 - dielectric 925
 - energy 1180
 - energy density 1157
 - parameter 1207
 - strength 89
- annealing 126, 1334
- anodic oxide film 678
- anodisation 678
- anti-ferroelectric 197, 604
- antiferromagnet 90
- antiferromagnetic
 - film 1164
 - NiO 1165
 - Pd₃₀Pt₂₀Mn₅₀ 1165
 - synthetic 1166
 - insulator
 - La₂CuO₄ 1203
 - YBa₂Cu₃O₆ 1203
 - pinning films 1185

- antiferromagnetically coupled (AFC) 1173
- antiferromagnetically coupled media (AFM) 1334
- anti-reflection coating 692, 694, 696
- anti-site defect 1334
- apparent bandgap narrowing 487, 1334
- areal density 1184, 1334
- array package 1268, 1270
 - physical property of material 1272
- array photoconductive 860
- array photovoltaic 858
- arrays
 - long wavelength (LW) 869
 - medium wavelength (MW) 869
 - two-color 880
 - via-hole type structure 880
- Arrhenius relation 122, 1334
- Arrhenius temperature dependence
 - inorganic glasses 219
- Arrhenius–Néel model 1172
- arsenic-based materials 276
- as-deposited film 689
- a-Se 1132, 1133
- a-Si:H 140, 566
 - dangling-bond-defect 581
 - dangling-bond-defect density 589
 - device applications 590
 - device-grade 582
 - drift mobility 574
 - electrical conductivity 574
 - electronic material properties 1107
 - electronic transport properties 1107
 - growth 582, 585
 - growth process 581
 - hydrogenated 581
 - optical properties 571
 - sputtered 1114
 - structural models 566
 - TFT 1107
- a-SiN_x:H film 1108
- aspect ratio 1004
- astigmatism 353, 1334
- A-swirl 104, 112, 1334
- defects 102
- asymmetric-AC sputtering 675
- atom diffusion 121
- atomic force microscopy (AFM) 279, 414, 660, 1233

- atomic layer deposition (ALD) 285
- atomic layer-epitaxy (ALE) 325, 328
 - growth rate 333
- atomic mixing 374, 383
- atomic scattering factor 349
- atomic sputtering rate 672
- atomic vibration 391
- attenuated total reflection (ATR) 587, 1334
- Au wirebond 1273
- Auger
 - depth profiling 378
 - electron 374
 - electron emission 345
 - hardware 375
 - instrument calibration 374
 - lifetime 862
 - map 375
 - quantitative analysis 374
 - spectrum 374
- Auger electron spectroscopy (AES) 373, 374, 1334
 - surface sensitivity 374
- Auger recombination
 - nonradiative 901
 - p-type material 870
- aurivillius
 - crystal structure 606
- auto exhaust sensor 223
- auto-compensation 1334
- automotive electronics 538
- avalanche
 - breakdown 424, 914
 - generation 471
- avalanche photodiode (APD) 914
 - single-photon counting 1034
 - telecommunication 1034
- average grain size 32
- average hopping time 183
- Avrami index 402
- azo compound
 - electrophotographic 972
 - photosensitivity 973
 - pigment 972

B

- back end of line (BEOL) 1334
- background flux 861, 865
- background radiation 861
- background-limited
 - detectors (BLIP) 858
 - device performance 865
 - operation 862
 - performance 868

- backscattered electrons 1334
- backscattered electrons (BSE) 348
- Baldereschi concept 1334
- ballistic
 - conductance 1150
 - electron 38
 - electron emission 810
 - SWNT 1150
 - transport 1152
- ballistic electron emission
 - microscopy (BEEM) 148, 151, 1335
- band
 - diagram 634
 - formation 1026
 - mobility 1110
 - overlap 98
 - structure 500, 895
 - structure, spin-polarised 95
 - tail states 176, 177
 - theory 1221
- band alignment 1335
- type I 1024
- type II 1024
- band filling 1335
- band gap 442, 1125, 1335
 - direct 848
 - engineering 481, 1021, 1023, 1024, 1335
 - narrowing 59
 - voltage 1029
- band gap energy 54, 735
 - cubic Group III–V ternaries 745
 - Group III–V binaries 743, 745
 - Group III–V quaternaries 745
 - Group III–V ternaries 743
- band gap quasi-particle 154
- band gap structures
 - graded 124
- band offset 1335
- band-edge modulated films 578
- band-pass filter 696
- band-stop filter 696
- bandtailing 59
- band-to-band absorption 57, 65
- band-to-band tunnelling 871, 912
- bandwidth 868
- barium strontium titanate (BST) 649, 1012
- Barkhausen noise 87
- barrier energy 201
- barrier height 148, 149, 1335
- basal plane dislocation 543
- batteries, high energy density 225
- BCS theory 1198
- beam deflection 1115
- beam effective pressure (BEP) 510
- beam effective pressure ratio 1335
- Beer–Lambert law 190
- Bell Laboratories 4, 5
- bend deformation 926, 927
- bend elastic constant 927
- beryllia, thick film technology 720
- bias sputtering 675
- biaxial elastic modulus 1108
- BiCMOS 7
- BICUVOX 225
- BIMEVOX 219
- bimolecular reaction 1100
- binary alloy, solid solution 26
- binary endpoint 741
- binary parameters 736
- biocompatible surface 1232
- bioluminescence 984
- bio-sensors, nanotube based 1153
- bipolar transistor 6, 430, 481
- bipolar transport 1030
- birefringence 932, 944
- birefringent crystals 1335
- bisazo compound 972
- bis-polycarbonate (Lexan) 180
- BJT (bipolar junction transistor, SiC) 547
- black-body
 - emission 983
 - radiation 892
- BLIP 1335
- Bloch wave vector 1026
- Bloch waves 20
- blocking contact 1127
- blocking layer 969
- blocking temperature 1335
- blue LED 287
- blue-violet laser 1144
- Blu-Ray system 12, 753
- Boltzmann constant 58
- Boltzmann equation (BE) 1335
- Boltzmann transport equation 29, 37, 447, 806, 807
- bombarding particles primary 383
- bond failure 1272
- bond switching 576
- boron-doped diamond 557
- borosilicate
 - crown 76
 - glass 1065
- Bose–Einstein distribution 58
- boule 1335
- boundary effects 412
- bowing parameters 743
 - Group III–V ternaries 743
- Bragg equation 348
- Bragg grating 998
- Bragg's law 1335
- branch-point energy 154
- Bravais lattice 1335
- Bremsstrahlung 1335
- Bridgman 303, 337
 - crystals 306
 - growth 247, 250, 336, 1335
 - method 236
 - process 305
- Brillouin zone (BZ) 441, 500, 809, 891, 913, 987, 998, 1335
- broadband sensitizer 1336
- broken bonds 1339
- Brooks–Herring (BH) formula 1336
- B-swirl 105
- buckminsterfullerene 1225
- bucky materials 1153
- buffer chambers 317
- buffer layer 754
- bulk
 - concentration 221
 - crystal growth, vapor phase 333
 - crystal structure 682
 - defects 430, 434
 - electron 696
 - films, dielectric constants 630
 - Group-III nitrides 776
 - growth 231, 245, 247, 304
 - heterostructure device 907
 - hole concentration 419
 - material conductivity 696
 - modulus 740, 757
 - resistivity 410, 411
 - silicon 101
 - single crystals 231
 - SiO₂ 636
 - SiO₂, point defects 638
 - valence electrons 154
- bulk AlN
 - thermal conductivity 763
- bulk modulus 1336
- bulk semiconductor 10, 1027, 1087
 - electron transport 807
 - nonlinear 1081
 - nonlinear response 1080
 - spreading resistance 414
- bulk wurtzite GaN
 - band structure 810
 - material parameters 809
 - valley occupancy 812
 - velocity-field characteristic 811
- bulk ZnO
 - oxygen vacancies 851
 - visible luminescence 851
- bulk-grown crystals 241

bulk-limited conduction 703, 705
 Burgers vector 357
 buried junction 877, 1099
 BX lines 851

C

C_{60} 1148
 C_{60} molecule 1225
 cadmium chalcogenide 685
 cadmium compound film 700
 – morphological characteristics 691
 cadmium mercury telluride (CdHgTe) 304
 calamitic 1336
 calcination 610
 cap layer growth 308
 capacitance
 – flat-band 425
 – method 422, 425
 – silicon surface 425
 – transducer 1254
 capacitance equivalent thickness (CET) 1336
 capacitance–voltage (C – V) 1336
 – technique 1336
 capacitance–voltage measurements 421
 capacitance–voltage–frequency (C – V – f) characteristic 1110
 capacitor dielectric 602, 629, 647
 – amorphous tantalum pentoxide 649
 – barium strontium titanate 649
 – DRAM 648
 – ferroelectric materials 650
 – non-volatile memory applications 650
 – scaling 648
 – thick film 723
 capping layers 1144
 carbon acceptors 505
 carbon atom
 – electronic states 1220
 carbon nanotube 1147, 1225
 – catalysts 1149
 – dimensionality 1148
 – electronic structure 1149
 – field-emission displays 1152
 – interconnect 1152
 – semiconductor 1149
 – symmetries 1149
 carbon-based
 – material 1220
 – molecule 1147

carbon-doped material 773
 Carlson, Chester F. 954
 Carnot efficiency 1195
 carrier
 – capture 1031
 – concentration 701, 767, 1336
 – concentration, direct determination 420
 – confinement 894, 898
 – density 214, 1110
 – distribution function 447
 – distributions 143
 – drift mobility 572
 – drift mobility calculation 183
 – effective mass 744
 – lifetime 311, 870
 – relaxation 1031
 – scattering 750, 767
 – temperatures 37
 – transport 465, 1031
 – tunneling 416
 carrier mobility 632, 744, 1336
 – diffusion picture 181
 – effective 1112
 – modulation 1113
 – nanotube 1148
 – time-dependent 181
 carrier range 1336
 carrier velocity
 – microscopic 1336
 carrier–carrier scattering 449
 carrier–lattice scattering 448
 carrier-to-noise ratio (CNR) 1145
 cascaded second-order materials 1336
 cast recrystallise anneal (CRA) 245, 305
 catastrophic breakdown 640
 cathodoluminescence 1336
 cathodoluminescence (CL) 343, 362, 367, 788, 984
 cation 1336
 cation self-diffusion 132
 cavitand compound 1259
 cavitand compound 1241
 Cd_3As_2
 – morphological characteristics 690
 Cd-based compound semiconductor 1336
 CdHgTe growth 279
 Cd-rich film 667
 CdS film 689, 708
 CdSe film 702, 703

CdTe (cadmium telluride)
 – band gap 1102
 – detector 1128
 – epilayer growth orientation 316
 – polycrystalline thin-films 1102
 – solar cell 1102
 – thin-film photovoltaic 1103
 CdTe-based materials 304
 CdZnTe 1124
 CdZnTe CZT 1128
 centro-symmetric 1336
 ceramic 1000
 – actuator, piezoelectric 617
 – capacitor 603
 – fabrication 610
 – laser ablation 615
 – materials, thick film technology 720
 – q-DC behaviour 209
 cermet (ceramic/metallics) 718
 cermet thick film 722
 – resistors 728
 cermets 1336
 chalcogenide glasses 139, 565, 569, 1336
 – band gap 573
 – dark conductivities 573, 575
 – drift mobilities 573
 – optical properties 572
 – PL spectra 572
 chalcogenides 401
 – Te-based 1139
 chalcopyrite phase $CuInSe_2$ (CIS) 1104
 channel mobility 550
 characteristic X-ray 1336
 characterization techniques 11, 410
 charge carrier
 – drift mobility 182
 – effective masses 1207
 – mobility 164, 166, 1222
 – relaxation kinetic 182
 charge carriers 1336
 charge coupled devices (CCDs) 410, 991
 charge generation 963, 1336
 – amorphous material 964
 – impact ionization 471
 charge pumping 1337
 charge pumping (CP) 428
 charge transfer complex 1150
 charge transfer interaction 1224
 charge transport 161, 180, 958, 1337
 – disordered materials 161
 – localized states 171

- charge transport layer (CTL) 956, 969, 974, 977
 - glass-transition temperature 975
 - organic photoreceptor 974
- charge-blocking layer 977, 1336
- charge-carrier 1202
 - lifetimes 1126
 - trapping 1136
- charge-coupled device (CCD) 839, 858, 878
- charge-generation layer (CGL) 956, 969, 970, 972, 977, 1337
 - bisazo compound 972
 - material 970
 - perylene 973
 - photoreceptor 970
 - phthalocyanine 973
 - trisazo compound 972
- charge-generation material (CGM) 1337
- charge-transport layer (CTL) 1337
- charge-transport material (CTM) 1337
- charge-transport phenomena 470
- chemical
 - electronegativity 154
 - microanalysis 350
 - self-diffusion 123, 132
 - sensing 1228, 1242
 - shift 375
- chemical annealing 586, 1337
- chemical bath deposition (CBD) 1103
- chemical beam epitaxy (CBE) 295, 359
- chemical sensor 1219
 - polymer 1255
 - thick film 730
- chemical solution deposition (CSD) 613
- chemical transformation 400
- chemical vapor deposition (CVD) 1337
- chemical vapour deposition (CVD) 257, 330, 546, 581, 660, 678
- chemical vapour transport (CVT) 246, 326
- chemically assisted ion beam etching (CAIBE) 363
- chemiluminescence 984
- chemiresistor 1228
- chip-fabrication facility (FAB) 1230
- chiral
 - dopant 922
 - index 1225
 - nanotube
 - diameter 1149
 - nematics 921, 1337
 - smectic C phase 922
- chirality 921
- chloride growth 508
- chromatic aberration 353, 1337
- CIP (current-in-plane) 1163
- circuit
 - design, power consumption 7
 - element interconnection 626
 - response delay time 627
 - semiconductor 418
- cleave and stain 1337
- close-space sublimation (CSS) 1103
- CMOS 7
 - gate stack 632
 - soft error 1278
 - technology 627, 1278
- CMT 1337
 - advanced structure 318
 - alloy composition 318
 - array 315
 - carrier lifetime 311
 - composition 310
 - dislocation density 315
 - epitaxial growth technique 304
 - epitaxial layer 311
 - hybrid array 320
 - impurity segregation 307
 - infrared detector 319
 - layer 318
 - material characteristic 311
 - MBE technology 317, 319
 - MOVPE 312
 - phase diagram 304
 - pseudobinary alloy composition 314
 - Si 320
 - sliding boat growth 311
 - structure 318
 - substrate material 315
- CMT growth
 - Bridgman 305
 - Hg-rich melt 309
 - monitoring 317
 - solid state recrystallization (SSR) 305
 - Te-rich 309
 - traveling heater method) 306
- coating
 - anti-reflection 692, 694, 696
- coatings
 - conformal 728
- Coats–Redfern–Sestak plot 402
- cobalt–iron film 1163
- Co–Cr–Pt alloy 1178
- co-dopant 845
- coefficient of thermal expansion (CTE) 761, 1272, 1337
- coercive field 88, 599, 1158, 1178, 1337
- coercive squareness parameter (S^*) 1337
- coercivity 1163, 1169
- coherence length 1200, 1207, 1337
 - in-plane 1207
- cohesive energy 1337
- Cole–Cole function 193
- collinear probe 413
- color glass 1070
- columnar grain structure 689
- columnar phase 1337
 - rectangular 924
- columnar recombination 1126
- compact disc (CD) 11
- compact disc player 529
- complementary characterization 364
- complementary metal oxide semiconductor 1337
- complex 1337
- complex heat capacity 404
- complex heterostructures
 - carrier concentration 522
 - composition measurements 522
- complex perovskite 604
- compliance constants 739
- composite materials
 - ionic conductivity 220
- compositional uniformity 1337
- compound annual reduction rate (CARR) 626
- compound semiconductor 500, 680, 1338
 - crystal properties 326
 - growth techniques 243
 - MOCVD technique 331
 - quality of epilayers 331
- compressive strain 1026, 1338
- compton scattering 1338
- computed radiography (CR) 993
- computing power 626
- concentration doping of cuprate 1202
- concentration quenching 1338
- conducting polymer 1241, 1255
 - conductivity 1255
- conduction 42
 - electrons 395, 396

- mean free path 21
- metallic-type 1109
- quantized 42
- tail states 59
- conduction band (CB) 57, 890, 1221, 1338
 - amorphous semiconductors 566
 - deformation 747
 - density of states 566
 - effective mass 445
 - potential energy 168
- conduction mechanism
 - bulk-limited 700
 - electrode-limited 700
- conductive layer 1338
- conductive polymer 1223
- conductivity 19, 1338
 - anisotropic 1284
 - electrical 575, 696
 - film thickness 697
 - ionic 217
 - ionic-electronic 213
 - layer-by-layer 1150
 - oxygen ion 220
 - temperature dependence 176
- conductivity mass 750, 1338
 - Group III–V binaries 746
- conductivity mixture rule 34
- conductivity relaxation time 20
- conductor 1267, 1282
 - ionic 214
 - thick film 613
- configuration coordinate model 1338
- configuration entropy fluctuations 205
- confined level 1030
- confinement
 - barriers (CB) 1031
 - energy 1027
- confining potential 1030
- conformal coatings 728
- conjugated polymer 180, 961
 - crystalline packing 1224
- constant-photocurrent method (CPM) 141, 567
- contact resistance 415
 - direct measurement 417
 - model 417
 - semiconductor 415
- container-free LPE (CFLPE) 275
- continuous-charging Czochozalski (CCZ) 267
- continuum theory 927, 928
 - chiral nematic 928
- convective flux 108
- conventional DSC 400
 - total heat flow 404
- convergent beam electron diffraction (CBED) 355
- converse piezoelectric effect 1338
- conversion detector
 - poly-TlBr 1130
- conversion efficiency 1098, 1338
- cooldown detector 859
- Cooper pair 1197
 - coherence length 1200
 - collective wave function 1198
 - density 1200
- Cooper pair density 1213
- Cooper pairs 1338
- COP 1338
- copper thick films 722
- core-level photoemission spectroscopy (PES) 152
- corona charging 969
- corona discharge 1338
- corona-discharge poling 1069
- correlation factor (f) 1338
- cost per transistor 626
- Cotterell atmosphere 1338
- Coulomb
 - blockade 1151
 - gap 173
 - interaction 63, 1027
 - island 1151
 - potential 766
- coupled granular continuous (CGC) recording media 1338
- coupled granular/continuous (CGC) medium 1181
- coverage ratio 1245
- CPP (current-perpendicular-to-plane) 1163
- critical (transition) temperature 1338
- critical current 1198
 - density 1209
- critical field 1338
 - metallic superconductor 1197
- critical fluctuations 1338
- critical temperature 1194
 - epitaxial strain 1213
- critical thickness 1338
- critically resolved shear stress (CRSS) 504
- cross Kelvin resistor (CKR) 417
- cross-luminescence 995
- crosstalk 875
 - long range 879
- cryogenic encapsulation 857
- cryopanel 1338
- crystal
 - density 737
 - key parameters 399
 - neckingless method 268
 - originated particle (COP) 105
 - phonon concentration 387
 - properties 326
 - structure 682
- crystal clear project 994
- crystal field splitting 985
- crystal growth 1022
 - conditions 107
 - history 232
 - hydrothermal method 336
 - striations 264
 - transport agent 333
- crystalline enthalpy 399
- crystalline material 756
 - diffraction peaks 350
 - non-centrosymmetric 598
 - properties 597
- crystalline polymer
 - thermal conductivity 394
- crystalline quality 542
- crystalline semiconductor 63, 573
 - Hall effect 574
 - inorganic 1083
 - optical absorption 63
 - thermoelectric power 574
- crystalline silicon 570, 1095
 - band gap 1099
 - intrinsic mobility 458
 - phonon modes 447
 - room-temperature mobility 458
 - structural models 566
- crystallization
 - kinetics 403
 - temperature 385
- crystallographic
 - polarity 358
 - structure 349
- crystals
 - annual production rates 232
 - Debye heat capacity 386
 - ionic 51
- c-Si 568
- cubic anisotropy 89
- cubic III–V binary 747, 748
 - compliance constant 739
 - elastic stiffness 739
- cubic III–V ternary 745
- cubic structures 13
- CuO₂
 - buckling in 1204
 - planes 1202

cuprate semiconductor

- crystal structures 1204
- CuO_2 planes 1202
- lattice parameters 1204
- cuprate superconductor
 - 1193, 1216
 - carrier concentration 1211
 - common features 1202
 - critical current density 1210
 - critical temperature 1205
 - electrical resistivity 1202
 - energy gap 1204
 - film 1213
 - grain boundaries 1211
 - hole concentration 1205
 - in-plane strain 1214
 - length scale 1207
 - upper critical fields 1207
- Curie temperature 84, 197, 599
- current in the plane (CIP) 1338
- current noise measurement
 - high-impedance devices 434
- current perpendicular to the plane (CPP) 1339
- current–voltage characteristic 1199
- current–voltage measurements 426
- cut-off wavelength
 - HgCdTe 862
- C – V measurement, low-frequency 425
- CZ crystals 1339
- Czochralski (CZ) crystal
 - growth conditions 263
 - seed-end portion 263
- Czochralski (CZ) growth 233, 250
- Czochralski (CZ) method 5, 9, 102, 249, 259, 451, 504
- Czochralski (CZ) pulling procedure 262
- Czochralski (CZ) silicon 264
 - carbon 265
 - crystal 263
 - crystal growth sequence 262
 - crystal impurities 264
 - dislocation-free 268
 - doping 265
 - impurities 266
 - oxygen 265
 - properties 261
 - striations 265
 - swirl 265
- Czochralski and liquid encapsulated Czochralski (LEC) growth 1339

D

- dangling bond 588, 638, 1339
 - a-Si:H 589
 - defect 581, 589
 - light-induced creation 576
- dark conductivity 137, 178, 576
- dark current 875, 1124
- dark decay 965, 1339
 - characteristic 965
 - photoreceptor 965
 - rate 957
- data rate 1162, 1184, 1339
- Davydov splitting 1247
- DC conductivity
 - temperature dependence 166
- DC hopping 706
- DC sputtering 614
- D-defects 105
- Debye frequency 386, 1198
- Debye heat capacity 386
- Debye length 1243, 1339
- Debye molar heat capacity 387
- Debye response 193, 207
- Debye temperature 24, 387, 388, 763, 764
 - III–V binaries 741
 - phonon concentration 393
- Debye theory 385, 387
- deep localized states 178
- deep traps 1125
- deep-energy-level impurities 1339
- deep-level transient spectroscopy (DLTS) 426, 520, 567, 1110, 1339
 - averaging techniques 435
 - electrical characterization 434
- deep-trapping time 1132
- deep-well drilling 538
- defect
 - analysis 103
 - chemistry 1339
 - defect interactions 220
 - density 106
 - GaAs 514
 - reaction 215
 - sources 367
 - types 106
- defective solids 220
- defect-related transitions in GaN
 - optical properties 785
- deformation potential 746, 1339
- deformation potential parameters 1339
- deformational phonon 448
- degradation 847
- delocalised states 164
- delta-doping 845, 849
- demagnetizing energy 90, 1170, 1172
- dense wavelength-division multiplexing (DWDM) 905
- density of state (DOS) 25, 42, 59, 138, 163, 395, 442, 746, 1109, 1110, 1339
 - 3-D 893
 - conduction band tail 1110
 - joint 1026
 - mass 1339
 - normal state 1198
 - single-electron 1198
 - spin-polarized 98
 - vibrational 386
- density of tail states 1111
- density-functional theory 159
- density-of-states 746
- deoxyribonucleic acid (DNA)
 - electron transfer 1236
 - electronic behaviour 1235
- depletion 1339
- depletion layer 150, 421, 1211
 - doping profile 422
- depletion region 701
- deposition
 - parameter 687
 - rate 667
- deposition method 659
 - chemical deposition 677
 - physical deposition 661
 - thin film 661
 - vacuum evaporation 662
- depth of field 354, 1339
- depth of focus 354, 1339
- depth profiling 1339
- depth resolution 1340
- detection efficiency 1340
- detection phase sensitive 991
- detective quantum efficiency (DQE) 1123, 1131, 1133, 1340
- detectivity 865, 1340
 - background-limited 868
- detector 1022
 - Hg – Te 866
 - $10\mu\text{m}$ 866
 - a-Se 1132
 - background-limited performance 868
 - cooldown time 868
 - detectivity 863
 - device parameters 861

- diffusion-limited 315
- exposure level 1131
- indium antimonide (InSb) 856
- infrared 855
- multiple quantum well (MQW) 857
- photoconductive 857, 858, 1036
- photovoltaic 868
- polycrystalline PbS 856
- poly-CZT 1132
- poly-HgI₂ 1132
- quantum efficiency 1133
- semimetals 866
- technology 855
- two-color 880
- white noise current 868
- X-ray sensitivity 1131
- zero bias resistance 868
- developer 955
- device
 - architecture 1230
 - characteristics 625
 - degradation 125
 - density 1230
 - dimensions 626
 - functionality 102
 - high-speed 9, 10
 - industry 101
 - optical 1068, 1076
 - parameters 861
 - performance 581
 - responsivity 861
 - structure 125, 755, 876, 1235
 - technology 121
- device photovoltaic 731, 868, 1095
- device-grade $\mu\text{c-Si:H}$ 581
- device-grade a-Si:H 581
- devitrification 1340
- DFB lasers
 - refractive index 905
 - threshold currents 905
- DFT (density-functional theory) 1340
- diamagnets 91
- diamond 537, 551
 - device fabrication 553
 - doping 551
 - emission properties 556
 - etchant 545
 - grown by CVD 540
 - heteroepitaxial 545
 - heteroepitaxy 546
 - high-temperature electronic devices 551
 - homoepitaxial 545
 - homoepitaxial growth 555
 - impurity 544
 - lattice 500
 - oxidation 545
 - p-channel FET 552
 - properties 544
 - resistor 557
 - single crystals 545
- diamond film 139, 545, 547
 - boron-doped 557
 - dielectric 557
 - field emission 556
 - homoepitaxial growth 555
 - nanocrystalline 557
 - polycrystalline 554
- diamond films phosphorus-doped 540
- diamond structure 1340
- diamond-like
 - carbon (DLC) 556
 - lattice structure 482
 - silicon 545
- dichlorosilane (SiH_2Cl_2) 257
- Dieke diagram 985
- dielectric 610, 625, 1267
 - contrast 1000
 - DC conductivity 208
 - dispersion 193
 - film 688
 - loss 199
 - materials properties 629
 - quantum mechanical tunneling 629
 - relaxation 202
 - reliability 640
 - response function 194
 - scaling 628
 - susceptibility 189
 - thick film 613
- dielectric constant 48, 401, 447, 545, 627
 - conductivity contribution 449
 - cubic III–V binaries 748
 - effective 32
 - optical 154, 748
 - substrate 720
 - wurtzite III–V binaries 748
- dielectric constant static 457
- dielectric constant/susceptibility (DC/DS) 1340
- dielectric increment 199, 203
 - dynamic scaling 203
 - temperature dependence 199
- dielectric material 626, 669, 1340
 - energy barrier 634
 - microelectronic devices 625
 - polarizability 629
- dielectric pastes
 - thick film 723
- dielectric permittivity relative 629
- dielectric permittivity 648, 925, 933
- dielectric relaxation 145
- dielectric response
 - basic features 190
 - basic macroscopic definitions 188
 - elements 209
 - low frequencies 208
 - physical concepts 187
- dielectric spacer layer 647
- dielectric susceptibility 1077
- dielectric thickness 648
- dielectric–semiconductor interface 640
- diethyl telluride (DETe) 313
- differential heat flow 400
- differential permeability 91
- differential scanning calorimeter (DSC) 385, 400
- differential scattering cross section 21
- diffraction 1340
 - contrast analysis 357
 - limit 1139
 - loss 1001
- diffusant
 - calculation 127
 - concentration 127
 - model 127
 - profiling 127
- diffusion
 - anneal 126
 - coefficient 126, 470
 - dopant 130
 - hot-carrier 470
 - isoconcentration 123
 - length 864
 - mechanisms 122
 - parameter 512
 - recombination enhanced 125
 - short circuit path 125
 - source 126
 - tracer concentration flux 123
 - transient enhanced 124
- diffusion coefficient or diffusivity 1340
- diffusion length 1340
- diffusion-induced disorder 124
- diffusion-limited detectors 315
- diffused base transistor 6
- digital circuits 9

- digital packaging 1270, 1272
 - thermal performance 1283
- digital printer 955
- digital printing 1340
- digital versatile disc (DVD) 11
- digital versatile disk (DVD) 888, 1140
- dihydrogen phosphate (KDP) 606
- diisopropyltellurium (DIPTe) 289, 313
- dilute magnetic semiconductor (DMS) 248
- dilute nitride 906
- dimensional scaling
 - capacitor dielectric 648
- dimethyl cadmium (DMCd) 289, 314
- dimethylzinc (DMZn) 289
- diode 410, 725
 - quantum-well (QW) 1036
- diode laser 1032
- diode sputtering 673
- dip coating 968
- dipolar fluctuations 191
- dipole density fluctuation 196, 210
 - relaxation rate 199
- dipole disorder model 962
- dipole fluctuations 193, 197
- dipole glass 197, 203
- dipole moment 188
 - effective 198
- dipole oscillator model 55
- dipole relaxation 207
- dip-pen nanolithography (DPN) 1233
- dipping 1340
 - boat 273
 - cycle 681
 - growth 310
 - techniques 308
- direct alloy growth (DAG) 314
- direct band gap 11, 1023
 - material 913
 - semiconductor 288
- direct gap semiconductors 570
- direct photo CVD 581
- direct piezoelectric effect 1340
- director 919, 924, 933, 1340
- director distribution 919
- director orientation
 - optical properties 934
- director reorientation
 - threshold voltage 933
- disc-like molecules 923
- discotic 1340
- discotic liquid crystal 924
- discotic nematic 923
- disk drives 1155
 - high-capacity 1155
 - rotary actuators 1177
 - track density 1177
- dislocation 11, 13, 356, 683, 767
 - grown-in 872
- disorder
 - model 962, 975
 - potential 166
 - static 62
- disorder model 1340
- disordered conductor 217
- disordered material 161, 1113
 - charge transport 167
 - DC conductivity 166
 - electrical conductivity 169
 - electron mobility 176
 - electron transport 170
 - extended states 167
 - hopping charge transport 169
 - transport phenomena 163
- disordered organic material 180
 - carrier mobility 180
 - charge transport 180
 - hopping conductivity 180
- disordered semiconductor 1108
 - AC conductivity 164
 - density of band carrier 1112
 - energy spectrum 163
 - hopping conduction 169
 - localized states 169
 - materials 162
 - transport properties 164
- disordered solid 962
 - electrical conduction 169
- disordered sublattice 219
- disordered systems
 - doped crystals 162
 - transport phenomena 162
 - transport theory 165
- disperse red 1085
- dispersion 48
 - ionic crystals 51
 - relation 51, 52
 - semiconductors 51
 - semiempirical single oscillator 52
- display devices 925
- Dissado–Hill function 194
- distributed Bragg reflector (DBR) 905, 1033
- distributed feedback (DFB) 889
- distributed feedback (DFB) laser
 - active layer 904
 - guiding layer 904
- distribution function 1341
- ditertiarybutylselenide (DTBSe) 289
- DMOSFET 550
- DNA
 - chip 1236
 - computer 1236
 - electronics 1219
- domain walls
 - silicon carbide 543
- donor 1341
- donor diffusion
 - n-type GaAs 513
- donor–acceptor
 - charge-transfer 972
 - pair (DAP) 785
 - photoreceptor 972
- donor-bound exciton (DBE) 778, 782
- dopant 272, 1341
 - concentration 124, 126
 - density 216
 - diffusion 124, 125, 132
 - Group III, V 128
- doped polymer 1222
- doped semiconductor 1341
- doping 5
 - by diffusion 847
 - concentration 422, 815
 - profiling 424
 - properties 12
- DOS effective mass 442
- double heterojunction (DH) 530
- double heterostructure (DH) 11, 12
- double Schottky barriers 1211
- double-channel planar buried heterostructure (DCPBH) 899
- double-layer charging 1152
- double-layer heterojunction (DLHJ) 278, 308, 876
- double-layer planar heterostructure (DLPH) 877
- drain depletion region 430
- drain resistances 427
- DRAM 1341
- DRAM planar capacitor 649
- drift length 861
- drift measurement 451, 462
- drift mobility 20, 1125, 1341
 - concentration dependence 184
 - temperature dependence 180
- drift velocity 8
 - electric field relationship 467
- drift-diffusion relation 446
- Drude
 - approximation 477

- formula 450, 1341
- model 55
- drum photoreceptor 1341
- insulating material 969
- dry etching 1004
- drying 610
- DSC cell
 - heat-flux 400
- dual layer organic photoconductor 1341
- dual-beam photoconductivity (DBP) 141
- dual-color
 - device 880
 - technologies 881
- dual-in-line package 725, 1268
- dual-layer disks 1144
- Dulong–Petit rule 387
- DVD 565
 - phase-change 1140
 - RAM 1140
 - recording 1140
 - RW 1140
- d-wave symmetry 1203
- DX centers 845
- dynamic coercivity 1341
- dynamic random-access memory (DRAM) 616, 647, 1229
- dynamic resistance 879
- dynamic SIMS (DSIMS) 377, 378

E

- E* layer 1174, 1341
- E'* center 1341
- easy axis 1341
- ECA metal particles 1282
- ECB cell
 - transmission curve 933
- edge filter 696
- edge-defined film-fed growth (EFG) 249
- edge-defined film-growth technique 1341
- edge-emitting laser 1021
 - elliptical beam profile 905
- effective dielectric constant 32
- effective dipole moment 198
- effective hole mass 746
- effective magnetic field 1187
- effective mass 1341
- effective media approximation (EMA) 32
- effective mobility 1112, 1113
- effective resistivity 23
- effective scattering 21

- efficiency 1098
- effusion cell 293, 509
- Einstein relation 1341
- EL(2) 1341
- elastic
 - constants 926
 - energy density 926
 - modulus 401
 - stiffness 739
- elastic compliance constant (*C*) 1342
- elastic deformation 1342
- elastic properties 739
 - semiconductor alloys 741
- elastic stiffness constant (*S*) 1342
- elasticity 1342
- elastoresistance 1113, 1114
 - coefficient 1114
- electric torque 931
- electrical conductivity 222, 575, 696
 - amorphous semiconductors 573
 - metallic film 696
 - nanotube 1148
 - silicon 452
 - size-dependent 698
 - Temperature-dependence 574
- electrical energy 1095
- electrical measurement techniques 410
- electrical properties 565, 696
 - aluminium nitride 773
 - bulk SiO₂ 638
 - chalcogenide glasses 574
 - electrical conductivity 574
 - gallium nitride 767
 - Groupx III nitrides 766
 - Hall coefficient of a-Si:H 574
 - indium nitride 774
 - thin film 660
 - thin films 696
- electrical pumping 890
- electrical resistivity 27
- electrically active impurities
 - point defects 434
 - SI GaAs 515
- electrically conductive adhesives (ECA) 1281, 1342
- electrically controlled birefringence 932
- electroabsorption 69, 1342
- electroabsorption modulator (EAM) 1037
- electroactive compound 1224

- electrochemical etching 1005
- electrochemical potential 1342
- electrochemical recognition 1231
- electrochromic elements 226
- electrochromic windows 223, 226
- electrode-limited conduction 706
- electrodeposition 677, 831
- electroding 612
- electrodynamics 1197
- electroluminescence 788, 984
- electrolytes
 - solid oxide 215
- electrolytic domain 222
 - boundaries 222
- electromagnetic (EM) 55
- electromechanical coupling 618
- electron 1342
- electron affinity 1342
- electron beam induced conductivity (EBIC) 343, 362
- electron beam lithography 1342
- electron beam–material interaction 346
- electron cyclotron PECVD 680
- electron cyclotron resonance (ECR) 680
- electron device
 - high-frequency 805
- electron diffraction 351, 353, 356, 359
- electron drift velocity 466, 813
 - AlN 820
 - doping dependence 816
 - GaAs 820
 - GaN 820
 - InN 820
- electron effective mass
 - density-of-states 746
 - III–V quaternaries 746
- electron energy analyser 375
- electron energy loss spectrometry (EELS) 343, 346, 367
- electron generation 1133
- electron injection 700
- electron mobility 166, 766, 767, 774, 1125
- electron motion 20
- electron nuclear double resonance (ENDOR) 850
- electron paramagnetic resonance (EPR) 513, 639, 850
- electron plane wave 1186
- electron saturated drift velocity 470

- electron scattering 872
- electron schubweg 1131
- electron spectroscopy for chemical analysis (ESCA) 343
- electron spectrum 443
- electron spin resonance (ESR) 567, 593, 639
- electron system 1222
- electron transport 805–807, 824
 - AlN 817
 - InN 819
 - material parameter 808
- electron wavefunction 1067
- electron-beam evaporation 666
- electron-beam lithography (EBL) 1001, 1003
- electron-beam poling 1069
- electron–cyclotron resonance (ECR) 754
- electroneutrality equation 221
- electron–hole overlap 1028
- electron–hole pair (EHP) 913, 1027, 1123, 1131
 - reduced mass 1027
- electronic
 - diamond 551
- electronic conductivity
 - mixed ionic 224
- electronic device
 - diamond-based 545
 - high-temperature applications 547
 - polymeric organic compound 1225
 - silicon carbide 547
- electronic material 805, 1267
 - C₆₀ based 1153
 - Debye heat capacity 388
 - heat capacity 388
 - silicon dioxide 635
 - thermal conductivity 391
 - thermal properties 385
- electronic noise 1134
- electronic nose system
 - sensor 1261
- electronic optical nose 1255
- electronic packaging 1267
 - classes of material 1269
 - encapsulation 1280
 - interconnect 1274
 - levels 1268
 - material 1267, 1269, 1272
 - solder 1274
- electronic transport devices 735
- electronics 1219
- electron–lattice 987
- electron–nuclear double resonance (ENDOR) 567
- electron–phonon
 - coupling 572, 573
 - interaction 448, 450, 1198
 - inter-valley scattering 463
 - scattering 1342
- electron-transport material
 - charge mobility 975
- electro-optic coefficient 604
- electro-optic device 608, 931
- electro-optic effect 1012, 1078, 1342
- electrophotographic cycle
 - printing 967
 - process parameter 958
- electrophotography 953, 1342
- electroplated alloys 1161
- electroplated Ni₄₅Fe₅₀ 1160
- electroplating 677, 1159, 1342
- electrospinning 1259
- electrostatic flux density 188
- elemental maps 375
- Eley–Rideal reaction 1342
- ellipsometry 72, 1342
- Elovich model 1251
- emergency lighting 993
- emission device 1342
- emissive transition 1028
- emitter 1032
 - interband 1032
- encapsulation 1342
 - electronic package 1280
 - material 1280
 - mold compound 1281
 - processing 1281
 - temperature 1281
 - wire sweep 1281
- end-member perovskite 602
- endothermic glass-transformation enthalpy 400
- endpoint materials 750
- energy band diagram 416
- energy conversion 224
- energy dispersive X-ray analysis (EDX) 343
- energy dispersive X-ray diffraction (EDX) 758
- energy gap 1335, 1342
 - (superconducting) 1198, 1203
- energy per carrier 1097
- energy relaxation
 - charge carriers 180
- energy transfer coefficient 669
- energy-band spectrum 1342
- energy-filtered TEM 12
- energy-loss relaxation 178
- enthalpy 386, 398, 401
 - crystalline 399
 - exothermic 400
 - fusion 398, 399
 - relaxation 399
- epi wafer 103
- epilayer
 - ellipsometry 313
 - growth 895
 - laser reflectometry 313
 - lattice constant 894
 - material 894
 - monitoring 313
- epitaxial CdTe layer
 - morphology 690
- epitaxial film 754
 - critical current densities 1212
- epitaxial GaAs
 - point defects 516
- epitaxial growth 684
 - techniques 231
- epitaxial layer (or Epilayer) 1342
- epitaxial layer overgrowth (ELOG) 13, 273, 277, 754
- epitaxial strain effects 1214
- epitaxies 271
- epitaxy 1342
- equation of state (EOS) 758
- equilibrium band bending 40
- equilibrium population 198
- equipartition of energy 387
- equivalent oxide thickness 1343
- erbium-doped fiber amplifier (EDFA) 907
- Er-doped AlN 787
- error rate 1170
- ESCALAB chamber 317
- etch pit density (EPD) 338, 1343
- etchant
 - diamond 545
 - film 545
- etching
 - electrochemical 1005
- evaporation
 - coefficient 663
 - of compounds 666
 - point source 664
 - process 665
 - reactive 667
 - source 664
 - surface source 664
- evaporator temperature 685
- Ewald sphere 1343
- Ewald sphere construction 354

EXAFS (extended X-ray absorption fine structure) 568
 excess noise 872
 exchange energy 93
 exchange field (H_{ex}) 1343
 exchange interaction 91
 exchange magnetic field 1174
 excimer laser annealing 845
 exciton 47, 1027, 1343
 – dipole moment 1070
 – energies GaN 781
 – ground state 1027
 – recombination 963
 – two-level system 1070
 excitonic absorption 65
 excitonic transitions 778
 exciton–phonon interaction 333
 exhaust sensor 223
 exhaustion concentration 1343
 exothermic enthalpy 400
 experimental methodology 370
 exponential DOS 1113
 extended states 163, 1109, 1110
 extended X-ray absorption fine structure (EXAFS) 1141
 extinction coefficient 48
 extraordinary refractive index – effective 933
 extrinsic doping 870
 extrinsic ferroelectric – Curie temperature 599

F

fabrication 976
 – nanoscale patterning method 1006
 – photoreceptor 976
 – technology 499
 Fabry–Perot 904
 – cavity 896
 – interference filter 696
 face-centered cubic (fcc) sublattices 755
 Faraday effect 69
 fast ion conductors (FIC) 216
 – structural characteristics 217
 fast switching displays 922
 fatigue
 – thermomechanical 1275
 Fermi
 – functions 891
 – level 148, 847, 892, 1343
 – surface 42
 – velocity 696

Fermi energy 25, 170, 445, 1109, 1198, 1343
 – density of states 1167, 1186
 Fermi's golden rule 1026
 Fermi–Dirac
 – distribution 152, 447, 891, 1028, 1111
 ferrimagnet 90
 ferroelastic 607, 1343
 ferroelectric 197, 1343
 – application 599
 – ceramic 198, 615
 – characteristic properties 599
 – Curie point 604
 – device electroding 612
 – dielectric response 203
 – domain 599, 1343
 – GeTe 1142
 – hysteresis 599, 1343
 – laser ablation 615
 – oxide 604, 614
 – oxide thin film 615
 – properties 598
 – sputtering 615
 – system 600
 – theoretical model 600
 – thin film 615
 ferroelectric application
 – capacitor 616
 – dielectric 616
 – piezoelectric 617
 – pyroelectric material 620
 ferroelectric ceramic
 – fabrication technology 610
 – poling 612
 ferroelectric material 650
 – aurivillius compound 606
 – Czochralski growth 608
 – fabrication 608–610
 – ferroelectric oxide 601
 – flux growth 609
 – illmenite 604
 – lead germanate ($\text{Pb}_5\text{Ge}_3\text{O}_{11}$) 607
 – oxide 606
 – Perovskite ferroelectric 601
 – phosphate 606
 – polycrystalline ceramic 610
 – polymeric 607
 – properties 597
 – reliability 651
 – single crystal 608
 – solution 609
 – thick-film 613
 – thin film 613

– triglycine sulphate 607
 – tungsten bronze 605
 ferroelectric random-access memory (FRAM) 617
 ferroelectric relaxor 1343
 ferroelectric to store this application
 – dynamic random-access memory 616
 ferromagnetic
 – configuration 1167
 – film 1164
 – layer 1173
 – material 83, 98
 – resonance (FMR) 88
 – solids 82
 ferromagnets 90
 fiber optic sensor 1228
 fiber optic window 1021
 Fick
 – diffusion 107
 – first law 122
 – law 512
 – second law 127
 Fick's first law of diffusion 1343
 Fick's second law of diffusion 1343
 field effect transistor (FET) 630, 735, 1151, 1226, 1343
 field emission 1343
 field emission characteristics
 – polycrystalline diamond films 557
 field emitter triode
 – monolithic diamond 555
 field factors 33
 field-effect mobility 1111
 field-emission display 1152
 figure of merit (FOM) 628, 1068, 1077, 1078, 1084, 1343
 – solar cells 1096
 – time-averaged nonlinear 1079
 fill factor (FF) 1343
 film
 – anodic oxide 678
 – antiferromagnetic 1164
 – as-deposited 689
 – a-SiN_x:H 1108
 – CdS 689, 708
 – CdSe 702, 703
 – cobalt–iron 1163
 – crystal structure 685
 – diamond 139, 547
 – epitaxial 754
 – ferromagnetic 1164
 – free 1185
 – in-plane strain 1214
 – insulating 700, 707
 – morphology 691

- oxide 674
- pinned 1164
- Poisson's ratios 1115
- thickness 661, 686, 689
- film growth
 - edge-defined 235
- film phthalocyanine 703
- film-growing surface 583
- filter
 - band-pass 696
 - band-stop 696
 - design 696
 - edge 696
 - semiconductor-doped 1070
- finite-well model 1025
- firing 1343
- first-order transition 925
- flash evaporation 668
- flash memory element
 - nitrided SiO₂ 649
- flat panel display 590
- flat-band capacitance 425
- flat-band voltage 435
- flatband voltage (V_{FB}) 1343
- flexible plastic substrates 1103
- flexible substrate 720, 1107, 1108
- flexoelectric coefficients 929
 - measurement 947
- flexoelectric effects 947
- flexoelectric polarisation 940
- flexoelectricity 928, 939, 1343
- flip chip 1270, 1344
 - device 1268
 - interconnect 1276
 - organic package 1278
 - solder deposition technique 1276
 - substrate 1278
 - technology 1278, 1280
 - underfill material 1280
- float zoning 6
- floating gate current 1344
- floating gate technique 427
 - semiconductor parameter 428
- floating zone technique 1344
- floating-zone (FZ) method 102, 112, 237, 250, 259, 451
 - properties 261
 - silicon 260, 261
- flow pattern defect (FPD) 105
- fluence 1344
- fluorescence 984, 991, 1344
 - quenching 965
- fluoride glasses 1344
- fluoroscopy 1124
- flux lines (vortices) 1344

- flux pinning 1344
- flux quantum 1199, 1200
- flying heads 1155
- flying height 1344
- focal plane array (FPA) 278, 305
 - large-area 304
- focused ion beam (FIB) milling 363
- focused ion-beam (FIB) lithography 1003
- forced convection 1344
- forward biasing 433
- Fourier transform infrared (FTIR) 311, 518
 - spectrometry 306
- Fourier's heat conduction law 391
- Fourier-transform infrared (FTIR) spectrometry 1344
- four-point probe 1344
- four-point-probe technique 410, 412
- Franck–Condon principle 990, 1344
- Frank–van der Merwe (FM) growth 332
- Franz-Keldysh effect 69, 478
- free carrier 47
 - absorption (FCA) 55, 56
 - density 140, 219
 - lifetimes 143
 - mobility 138
- free convection 1344
- free electron model 26
- free film 1185, 1344
- free-exciton transitions 781
- Frenkel
 - defect pairs 215
 - disorder 218, 221
 - excitons 64
 - formation energy 216
 - pair 122
- frequency (LF) 421
- frequency response 1223
- Fresnel coefficients 692
- Fresnel reflectance spectrum 472
- friction coefficient 545
- Friedel's law 1344
- front end of line (FEOL) 1344
- Fréedericksz transition 933
- fuel cell
 - monolithic 225
 - solid oxide 224
- fullerenes 1147, 1148
 - spherical-shaped 1225
 - transition temperature 1195
- fundamental properties 692

- fusion enthalpy 398, 399
- fusion temperature 399

G

- GaAs electro-optical devices
 - multi-quantum-well (MQW) laser 530
- GaAs_{1-x}P_x
 - fabrication 508
- GaCl-pretreated sapphire 768
- gain condition 899
- gain-guiding 898
- gallium arsenide (GaAs) 9–11, 358, 499, 503, 814, 820
 - band gap 1101
 - band structure 500
 - bulk growth 502
 - carbon concentration 518
 - carbon diffusion 513
 - carbon impurities 505
 - cathodoluminescence 521
 - chloride growth 508
 - conducting 520
 - Cr-doped 504
 - crystal defects 514
 - deep electronic levels 515
 - deep level defects 519
 - defect analysis 517
 - device 499
 - device-quality 295
 - diffusion characteristics 512
 - DLTS 520
 - doped 513
 - doping levels 510
 - electrical analysis of defects 518
 - electrical resistivity 501
 - epitaxial growth 507, 511
 - from the melt 506
 - homoepitaxy 754
 - hydride growth 508
 - impurity elements 517
 - lattice constants 501
 - liquid-phase epitaxy (LPE) 507
 - melt-grown 507, 515
 - MESFET 526
 - MOCVD growth 508
 - modulators 531
 - MOSFETs 10
 - n-type doping 510
 - optical analysis of defects 521
 - optical properties 514
 - phase diagram 502
 - photodetectors 532
 - Photoluminescence 521
 - Shallow Acceptors 513

- shallow donors 503, 513
- solar cells 1102
- surface quality 521
- transition metals 513
- vertical gradient freeze (VGF) 506
- wafers 505
- gallium arsenide (GaAs) crystal
 - phase diagram 502
- gallium arsenide (GaAs) device
 - electrical contacts 524
- gallium arsenide (GaAs)
 - electro-optical devices
 - emitters 527
 - GaAs photodetectors 532
 - heterojunction laser 529
 - lasers 529
 - modulators 531
 - solar cells 532
- gallium arsenide (GaAs) melt
 - horizontal Bridgman 503
 - horizontal gradient freeze technique 503
- gallium arsenide (GaAs)/GaAlAs laser 530
- gallium arsenide (GaAs)-based devices
 - precursors 509
- gallium nitride (GaN) 12, 13, 287, 350, 359, 537, 805, 814, 820
 - alloy 288
 - average electron energy 812
 - based devices 791
 - bound excitons 782
 - conduction band 778
 - donor-acceptor transitions 785
 - electrical properties 767
 - epilayer 779
 - exciton energies 781
 - fundamental properties 755
 - lattice constants 757
 - lattice parameter 762
 - layer 767
 - mechanical properties 759
 - optical phonon frequencies 760
 - optical properties 778, 785
 - template 754, 784
 - wurtzite 812
- galvanomagnetic effect 452
- galvanomagnetic measurement 451
- gap states 145
- Ga-rich GaAs
 - electrical properties 502
- garnets 92
- gas phase diffusion 592
- gas precursors 331
- gas sensor 1244
 - impedance-based paste 730
 - inorganic material 1243
 - potentiometric 223
 - semiconductor 1243
- gas source MBE (GSMBE) 295, 511
- gate capacitance 631, 1111
- gate dielectric 629, 630, 644
- gate dielectric material 641
 - reliability 635
- gate dielectric thickness scaling 628
- gate electrode work function 634
- gate insulator 158
- gate leakage currents 634
- gate length 1230
- gate oxide 8
- gate oxide integrity (GOI) 102, 1344
- gate-substrate capacitance 436
- Gaussian DOS 181
- Gay-Berne mesogen 919
- Ge₂Sb₂Te₅ GST 1140
- geminate recombination 964, 1126, 1344
- “GEN III” detector 880
- generalized Drude approximation (GDA) 449
- germanium 5
- getter sputtering 675
- ghosting 1345
- giant magnetic resonance (GMR) 1167
 - physical origin 1167
- giant magnetoresistive (GMR) 1162
 - head operation 1188
 - read sensor 1156
 - sensor 1157, 1175
- giant-magnetoresistive (GMR) effect 1345
- Gibbs free energy 199, 398
- Ginzburg–Landau parameters 1207
- Ginzburg–Landau theory 1200
- Gladstone–Dale coefficient 52
- glass 47, 1063
 - aging 404
 - band gap energy 1066
 - ceramics 1070
 - enthalpy 399
 - fast ionic conductivity 219
 - formation 202
 - glass-transformation kinetics 404
 - homogeneous 1065
 - key parameters 399
 - matrix 1071
 - metallic 396
 - nanoparticle dispersed 1070
 - optical properties 1065
 - oxide 52, 390
 - particle-embedded 1070
 - poling methods 1069
 - quantum-dot 1087
 - refractive index 1065
 - structure 401
 - substrate 1108
 - thermal conductivity 393
 - third-order nonlinearity 1065
 - transition region 404
- glass ceramics 1345
- glass frit 1345
- glass matrix
 - electrical conductivity 161
- glass surface
 - hydrophilic 1245
- glass transition 202, 401, 949
 - polymer 389
 - temperature 219, 385, 389, 399, 961, 975, 1281
- glasses
 - ionic transport 219
- glass-forming liquid 398
- glass-forming material 206
- glass-transformation kinetics 404
- glass-transition temperature 1345
- glassy metals 396
- glow curve 990
- glow discharge
 - mass spectrometry (GDMS) 376, 377, 1358
 - mass spectroscopy (GDMS) 838
 - optical emission spectroscopy (GDOES) 373
 - spectroscopy 373
 - spectroscopy (GDOES) 376
 - sputtering 673
- glow discharge spectrometry (GDS) 1345
- glow-discharge mass spectrometry (GDMS) 517
- gold conductors
 - thick film 721
- graded refractive index (GRIN) 530
- gradient freeze growth 1345
- gradient freezing (GF) 337
- grading 1345
- grain boundaries 30, 1129, 1210

- grain size 1127
 - distributions 1170
- grain-boundary
 - conductivity 699
 - defect 1127
 - scattering 698
- granular media 1182
- granular polysilicon 258
- graphite 1224
- graphitised SWNT 1149
- Gratzel cell 1345
- GRINSCH laser 530, 1345
- GRINSCH QW laser 531
- ground-state exciton 1027
- Group III-nitride
 - LEDs 13
- Group III–nitride system 538
- Group III–V 276
- Group III–V alloy
 - deformation potential 747
- Group III–V alloy systems
 - thermal conductivity 741
- Group III–V antimonides 276
- Group III–V binaries 745, 746, 750
 - crystal density 737
 - elastic properties 739
 - lattice parameter 737
- Group III–V compound 748
 - acceptors 503
 - diffusion 130
 - donors 503
 - growth 234
- Group III–V growth
 - binary 276
 - ternary 276
- Group III–V material 9, 1021
 - MOCVD 286
- Group III–V nitride semiconductor 805, 811
 - material parameter 808
 - recent development 824
 - transient electron transport 810
- Group III–V nitride semiconductor material 822
- Group III–V quaternaries 738, 745, 746, 750
 - lattice-matching condition 737
- Group III–V semiconducting compound alloy
 - electronic transport device 735
 - laser diode 735
 - light-emitting diode 735
 - optoelectronic device 735
 - photodetector 735
- Group III–V semiconductor 3, 499, 913
 - Group III–V ternaries 743, 750
 - Group III–V ternary
 - band-gap energy 743
 - optical mode 740
 - Group III–V ternary and quaternary alloy
 - optical mode 740
 - Group II–VI binary compound 131
 - Group II–VI compound 748
 - bulk single crystal 326
 - wide-bandgap 326
 - Group II–VI compound epilayers 330
 - Group II–VI compound semiconductor 335
 - bulk single crystals 326
 - high quality film 326
 - Group II–VI semiconductor 303
 - MOCVD 283
 - Group II–VI ternary compound 866
 - Group II–VI wide-bandgap compound
 - Bridgman 337
 - gradient freezing 337
 - group index 53
 - Group-III acceptors 445
 - Group-III nitride semiconductor 277, 753, 772
 - crystal structures 755
 - electrical properties 766
 - mechanical properties 757
 - specific heat 764
 - thermal conductivity 762
 - Group-IV semiconductor 128
 - Group-V acceptor 850
 - Group-V donors 445
 - growing crystal
 - axial temperature gradient 106
 - grown-in defects 106
 - growth
 - layer-by-layer 291
 - growth characteristics 873
 - growth method 235, 873
 - Bridgman 236
 - Czochralski 233
 - edge-defined 235
 - float zone 237
 - high-temperature solution 238
 - horizontal Bridgman 503
 - hydrothermal 239
 - Kyropoulos 235
 - large diameter 235
 - low-temperature solution 238
 - materials 240
 - Stepanov 235
 - travelling heater method 237
 - vapor 240
 - Verneuil high melting point materials 233
 - vertical gradient freeze 237
- growth model
 - μ c-Si:H 586
- growth process 108
- growth rate 249
- growth site formation 583
- Grüneisen
 - law 396
 - parameter 397
- GST
 - amorphous state 1142
 - bond lengths 1142
 - crystal structure 1141
 - ferroelectric 1145
 - local structure 1142
 - metastable crystalline 1145
- Gunn
 - diode 525, 1345
 - effect 501

H

- Hagen–Rubens relationship 56
- half-metals 98
- Hall
 - carrier concentration 420
 - coefficient 446
 - coefficient n-Si 459
 - conductivity 768
 - constant 164
 - factor 37, 446
 - measurements 418
 - mobility 36, 446, 751, 766
 - mobility III–V binaries 750
 - resistivity 43
 - scattering factor 420
- Hall effect 4, 19, 35, 36, 766, 1345
 - ambipolar conduction 36
 - carrier concentration 419
 - coefficient 36
 - materials characterization 418
 - measurements 767
 - physical principles 419
- Hall–Petch relationship 1273
- Hamiltonian
 - vibrational 987
- hand-held display 937
- hard axis 1345
- hard magnetic bias film 1345
- hardness 545, 758
- Havriliak–Negami function 193
- head field 1177

- head field slope parameter (Q) 1345
- heat capacity 1345
 - composition dependence 390
 - glasses 390
 - low-temperature experiments 387
 - normalized 390
- heat exchanger method 250
- heat flow 392
- heat-affected zone (HAZ) 1273
- heavy-hole
 - band 895, 909
 - effective mass 443
 - spectrum 468
- Heisenberg exchange 86
- helical anisotropy 89
- hermetic packaging 728
- herringbone structure 1248
- Hervé-Vandamme relationship 51
- hetero epitaxy 272, 281, 329, 356
- heteroepitaxial diamond 545
- heteroepitaxial layers 784
- heteroepitaxy 1346
- heterofullerides 1153
- heterogenous nucleation 402
- heterojunction bipolar transistor (HBT) 296, 482, 513, 526, 735, 1346
- heterojunction contact 862
- heterojunction laser 1346
- heterojunction structures 754
- heterojunctions 1346
- heteropassivation 871
- heterostructure 9
 - detector 311
 - lattice-matched 156, 157
 - metamorphic 157
 - non polar 157
 - pseudomorphic 157
 - waveguide 1001
- heterostructures 1346
- hexagonal
 - close-packed (hcp) 755
 - columnar phase 923
 - crystals 758
 - structure 13
 - wurtzite structure 756
- hexagonal As 1346
- HgCdTe 857, 878
 - 2-D arrays 879
 - array technology 870
 - Auger recombination 870
 - BLIP-limited performance 867
 - defects 872
 - diffusion current 870
 - diode 879
 - dual-band array 880
 - epitaxial growth 873
 - growth process 860
 - high-performance infrared systems 882
 - infrared detectors 860
 - loophole arrays 872
 - photoconductive 857
 - photoconductor array 860
 - photodiode (I–V) characteristics 869
 - photovoltaic array 867, 882
 - photovoltaic detector 882
 - slab 860
 - solution 310
 - spatial resolution 879
- HgI₂ 1124, 1127
- HgMnTe 1346
- high carrier mobility 545
- high electron mobility transistor (HEMT) 10, 297, 735
- high frequency (HF) 421
- high frequency device 549
- high performance FET fabrication 553
- high power SIT 549
- high refractive index glasses 76
- high temperature electronics 539, 544
- high vacuum (HV) 661
- “high- κ ” dielectrics 643
- high-angle annular dark field (HAADF) 365
- high-brightness diodes 287
- high-density recording 1180
- high-electron-mobility transistor (HEMT) 526, 1346
- high-energy sputtering 671
- higher operating temperature (HOT) 881
- higher silane-related species (HSRS) 592
- highest occupied molecular orbital (HOMO) 1226
- high-field conduction 700, 708
- high-frequency (HF) capacitance method 421
- high-frequency electron device 805
- high-index contrast 1346
- highly doped Si
 - room-temperature resistivity 454
- high-mobility InGaN 793
- high-order Laue zone (HOLZ) 355
- high-performance FET
 - DC output characteristic 553
- high-pressure depletion (HPD) 592
- high-purity n-Si 470
- high-quality AlN 787
- high-resistivity crystal 461
- high-resolution
 - displays 937
 - mask 1002
 - X-ray diffraction (HRXRD) 127, 756, 846
- high-speed device 9, 10
- high-temperature CVD (HTCVD)
 - silicon carbide 542
- high-temperature device 555
 - pn-diodes 558
 - Schottky diode 558
- high-temperature electronic
 - SiC-based JFETs 550
- high-temperature solution growth 238, 251, 1346
- high-temperature superconductor 1193
 - irreversibility line 1208
 - physical property 1195
 - upper critical field 1208
- hole
 - array 1004
 - drift velocity 466
 - effective mass 746
 - generation 1133
 - lifetime 1126
 - mobility 1125
 - spectrum 443
 - trapping sites 143
- hole transport
 - trap-free 962
- holographic patterning 1011
- homeotropic alignment 1346
- HOMO – highest occupied molecular orbital 1346
- homoeptaxial diamond 545
- homoeptaxial GaN 366
 - microstructural defects 366
- homoeptaxial GaN 783
 - layers 784
- homoeptaxy 272, 281, 1346
- homogenous nucleation 402
- homojunction 898, 1346
 - device 870
- hopper coating 968
- hopping
 - band 1113
 - conduction 705
 - conductivity 171, 178
 - distance, time-dependence 175
 - energy relaxation 177, 181

- length 177
 - rate 170, 175
 - transition 169, 178
 - transport 174, 180, 960
 - hopping conduction 1346
 - horizontal Bridgman (HB) 503
 - GaAs 503
 - growth 503
 - horizontal directional solidification
 - crystallization (HDC) 250
 - horizontal gradient freeze (GF) 504
 - GaAs 503
 - growth 503
 - hot carrier phenomena 1346
 - hot-carrier diffusion 470
 - hot-carrier phenomena 466
 - hot-wall epitaxy (HWE) 325, 328
 - II–VI compound 330
 - hot-wire chemical vapor deposition (CVD) 1346
 - humidity sensor
 - thick-film 730
 - Huygens's principle 348, 1346
 - HWE growth chamber 331
 - hybrid circuits 718, 1346
 - substrates 720
 - hybrid thin film 1255
 - hybridization 1346
 - hydride growth method 508
 - hydride vapor-phase epitaxy (HVPE) 754, 764
 - hydrogen dilution ratio 583
 - hydrogen plasma 587
 - hydrogen sensor
 - titanium dioxide nanotube 1244
 - hydrogenated 581
 - amorphous carbon transport properties 162
 - amorphous germanium (a-Ge:H) 61
 - amorphous silicon (a-Si:H) 61
 - hydrophobic substrate 681
 - hydrothermal growth 250, 1346
 - hydrothermal techniques 336
 - hysteresis 83, 1347
 - loss 84
 - time-dependent 85
-
- I**
- IBM copier 972
 - IBM copier 1 955
 - IC technology 626
 - IDCA 1347
 - ideal contact 150
 - ideal detector 1133
 - performance 1131
 - ideality factor 149, 151, 1347
 - IGBT (insulated gate bipolar transistor) 551
 - illmenite
 - LiNbO₃ 604
 - LiTaO₃ 604
 - illumination
 - nonresonant 1084
 - image contrast analysis 356
 - image formation
 - photoreceptor 956
 - image lag 1121, 1347
 - imaginary susceptibility component 193
 - imaging detector 1121
 - impact ionisation 914
 - impact ionization 38, 471, 862, 871, 1347
 - IMPATT diode 471
 - improper (or extrinsic) ferroelectric 1347
 - impurity 1347
 - absorption 66
 - ionization energy 445
 - metallic 103
 - scattering 10, 448
 - solubility 264
 - species 674
 - in situ monitoring 313, 1347
 - in situ oxidation 1188
 - incandescent light bulbs 753
 - index changes
 - negative nonlinear 1083
 - nonlinear 1085
 - index of refraction
 - second-order 1064
 - index-guided structures 898
 - index-guiding 897
 - in-diffusion 131
 - indirect-gap material 890, 987
 - indium nitride (InN) 805
 - electrical properties 774
 - optical properties 789
 - indium tin oxide (ITO) 1012, 1226
 - induced anisotropy 1157
 - induced dipole moment density 191
 - inductance
 - linear 85
 - inelastic scattering 1347
 - infinite well
 - wave function 1027
 - infinite-well 1025
 - information technology 1244
 - infrared (IR) 303, 914
 - infrared camera 858
 - HgCdTe 868
 - infrared detector 608, 860
 - cooler 858
 - minimal cooling 882
 - infrared laser absorption
 - spectroscopy (IRLAS) 582, 1347
 - infrared reflection absorption
 - spectroscopy (IRRAS) 584, 1347
 - InGaAsP 11
 - InGaN 8, 12
 - InGaN alloy 13
 - InGaN band gap 792
 - injection circuits 870
 - injection efficiency 869
 - ink-jet printing 1233
 - InN 820
 - electron transport 819
 - fundamental properties 755
 - lattice constants 757
 - material parameters 809
 - mechanical properties 759
 - optical phonon frequencies 760
 - Raman-active modes 760
 - inorganic crystalline semi-conductors
 - nonlinear properties 1080
 - inorganic glasses 219
 - in-plane coherence length 1207
 - in-plane transistor 553
 - in-plane wave 1027
 - insulating film 700, 707
 - insulating oxides 95
 - insulation layer
 - thick film 723
 - insulator 63, 1267
 - branch-point energy 159
 - dielectric band gap 154
 - dielectric constant 154
 - electrical characterization 410
 - resistivity 411
 - work function 700
 - insulator–metal transition 445
 - insulator–semiconductor interface 159
 - integrated circuit (IC) 6, 7, 725, 1347
 - gate dielectrics 629
 - manufacturing 626
 - nanometer scale 80
 - performance 626
 - scaling 626
 - wire bonding 726
 - integrated circuit electronics 5

- integrated electronic absorption 1065
 - integrated optics/photronics 1347
 - integration capacitance 878
 - intentional doping 773
 - interaction cross section 670
 - interatomic potential energy 397
 - interband emitter 1032
 - interband optical transitions 52
 - interband transition region 749
 - interconnect 1347
 - capacitance 628
 - dielectric materials 652
 - dielectric tetraethoxysilane (TEOS) 651
 - dielectrics 629, 651
 - electrical 1273
 - environment 1273
 - flip chip 1276
 - Pb-free 1277
 - wire material 627
 - interconnects
 - conductivity 1282
 - polymer 1282
 - second-level 1282
 - interdiffused multilayer process (IMP) 314, 1347
 - interdiffusion 131, 133
 - CMT 314
 - compositional 124
 - interface
 - band-bending 150
 - capacitance 209
 - dangling bond 639
 - electrical characterization 410
 - induced gap states (IFIGS) 147, 153
 - resistivity 416
 - trap capacitance 421
 - trap charges 424
 - trap density 1347
 - interface-induced gap states 1347
 - intergrain j_c 1347
 - intermediate diffuser 129
 - intermetallic compound (IC) 1194, 1274, 1275, 1347
 - intermetallic reactions and phases 1348
 - intermolecular forces 947
 - internal energy 387
 - internal photoemission yield spectroscopy (IPEYS) 148, 152
 - International Technology Roadmap for Semiconductors (ITRS) 626
 - interpolation
 - linear 747
 - interpolation scheme 736
 - interrupted field time-of-flight (IFTOF) 145
 - interstitial 1348
 - diffusion mechanism 122
 - nitrogen 847
 - oxygen 116
 - recombination 124
 - site 215
 - inter-sub-band (ISB)
 - emission 1030
 - transition 1029
 - transition rate 1030
 - intersubband (ISB) transition 1348
 - inter-valley electron transfer 776
 - inter-valley scattering 448
 - intragrain j_c 1348
 - intrinsic
 - magnetic properties 87
 - mobility 427, 452
 - optical transitions 777
 - pinning 1212
 - intrinsic point defect 107, 1348
 - aggregate void 104
 - aggregation 114, 115
 - Boron doping 114
 - Carbon doping 115
 - thermophysical properties 103
 - intrinsic resistivity 699
 - intrinsic silicon 128, 138
 - inversion 1348
 - inversion layer 421
 - inversion symmetry 1348
 - inverted opal 1010
 - inverted staggered TFT 1112
 - iodine doping 874
 - ion beam milling 874
 - ion bombardment 345, 593
 - ion conductor
 - disordered oxygen 219
 - ion implantation 874, 1003, 1348
 - GaAs 513
 - parameters 544
 - ion species primary 379
 - ion yield
 - positive secondary 380
 - relative 381
 - ion-beam-assisted deposition (IBAD) 760
 - ionic conduction 161
 - ionic conductivity 217
 - interface 220
 - ionic conductor 214
 - ionic conductors
 - application 214
 - ionic crystals 51
 - ionic disorder 216
 - ion-implanted SiC 548
 - ionization 1348
 - ionization energy 1348
 - ionized-impurity scattering 465
 - ion–lattice interactions 987
 - IPEYS technique (internal photoemission yield spectroscopy) 1348
 - IR detectors 869
 - irreversibility line 1209, 1348
 - irvin curve 1348
 - isoconcentration diffusion 123
 - isolation dielectrics 629
 - LOCOS isolation oxide 647
 - isothermal capacitance transient spectroscopy (ICTS) 567
 - isotope effect 1198
 - isotope labeling 588
 - isotropic liquid 919
 - itinerant magnet 93
 - I – V technique (current–voltage technique) 1347
-
- ## J
-
- JFETs (junction field-effect transistor)
 - SiC 547
 - Johnson–Mehl–Avrami equation 402
 - Josephson junction 1210
 - critical current 1200
 - Josephson–Fraunhofer interference 1200
 - Joule–Thompson cooler 857, 859
 - Judd–Ofelt (JO) analysis 68
 - jump process
 - diffusion coefficient 214
 - junction barrier Schottky (JBS) 548
 - junction characteristic 1098
 - junction geometry 876
-
- ## K
-
- $K_3Li_2Nb_5O_{12}$ (KLN) 249
 - Kelvin contact resistance (KCR) 410
 - Kerr effect 1011
 - Kevlar 918
 - Kikuchi lines 355, 1348
 - kinematic viscosity 949
 - kinetic theory of gases 663
 - Kirkwood factor 198

Knoop hardness 739
 Knudsen cell 509, 1348
 Knudsen effusion 293
 Kodak Ektaprint 100 955
 Kramers–Kronig relation (KKR)
 49, 195, 447, 1064, 1067, 1080
 Kramers–Kronig relations (KKR)
 1348
 Kretschmann’s configuration
 1252
 Kronig–Penney model 1026
 KTiOPO_4 (KTPO) 249
 Kyropoulos growth method 235,
 250, 339
 Kyropoulos technique 1348

L

Landauer formula 1150
 Landé factor 82
 langasite family ($\text{La}_3\text{Ga}_3\text{SiO}_{12}$) 249
 Langmuir–Blodgett (LB) 660, 1230
 – film deposition 680
 – method 1246
 – technique 677
 Langmuir–Blodgett film deposition
 1348
 Laporte’s rule 986, 1348
 laptop computer 937
 large area detector 1124
 large area electronic 1107
 large-band-gap material 766
 laser 11
 – ablation 615
 – diode 1032
 – edge-emitting 888
 – four-level 1034
 – quantum cascade (QC) 1033
 – spectral purity 904
 – vertical-cavity surface-emitting
 1032
 laser cavity
 – refractive index 897
 laser characteristics 900
 laser crystallization 1348
 laser detector 308
 laser diode (LD) 325, 735, 750,
 753, 888
 laser flash method 764
 laser light scattering (LLS) 102
 laser light scattering tomography
 defect (LSTD) 105
 laser trimming
 – thick film 726
 laser-induced fluorescence (LIF)
 582, 1349
 lateral conductivity 696
 lateral epitaxial overgrowth (LEO)
 754, 763
 lattice
 – constant 757, 1026
 – Group III–V binaries 737
 – melting 217
 – mismatch 315, 482, 1212
 – mobility 461
 – parameter 737, 756
 – reciprocal 354
 – vibrations 22
 lattice-matched $\text{AlInN}/\text{AlGaIn}$ 793
 lattice-matched system 95
 lattice-matching 1022, 1349
 Lattice-matching condition 738
 lattice-matching condition 737, 738
 law of mass action 1349
 law of mass action (LMA) 123
 layer
 – composition 968
 layer-by-layer growth 291
 LDA (local density approximation)
 97
 lead phthalocyanine film 705
 lead zirconate titanate (PZT) 729
 lead-free piezoelectric 603
 lead-silicate glass 1065
 leakage current 549, 630, 869, 872,
 902
 length between bits 1170
 lenses
 – growth 235
 Li batteries 223
 LIDAR (light detection and ranging)
 882
 lifetime 1349
 ligands 68
 light absorption 47
 light amplification 1068
 light emitting diode (LED) 3, 10,
 11, 325, 538, 735, 750, 753, 888,
 893, 1021, 1032
 – material properties 528
 – ultraviolet 1032
 – white 992
 light-hole
 – band 909
 – effective mass 443
 light-induced defect creation 1349
 light-induced phenomena 565
 – amorphous semiconductors 575
 lighting strip 992
 lightly doped drain (LDD) 415
 lineage 1349
 linear absorption 1067
 linear inductance 85
 linear interpolation 747
 linear interpolation scheme 736
 linear muffin-tin orbital (LMTO)
 758
 linear relaxation response 192
 linear susceptibility 195
 liquid
 – key parameters 399
 liquid crystal 917, 1012, 1349
 – alignment 925
 – bulk nematic 926
 – calamitic 919
 – chemical structure 942
 – chiral 921
 – chiral nematic 940
 – cholesteric phase 921
 – dielectric anisotropy 945
 – dipole moment 945
 – director 925
 – discotic 923
 – elastic constants 946
 – electrical conductivity 937, 943
 – helical structure 921, 945
 – long range order 924
 – lyotropic 918
 – material 941
 – molecular properties 943
 – molecular shape 919
 – molecular structure 919
 – negative dielectric anisotropy
 938
 – nematic 919
 – optical properties 932, 943
 – optical retardation 932
 – order 919
 – permittivity 945
 – positive dielectric anisotropy 925,
 932
 – rod-like molecules 924
 – smectic 919
 – thermotropic 918
 – torsional elasticity 947
 – viscosities 947
 – viscosity 929
 – viscosity coefficients 947
 liquid crystal shutter 940
 liquid encapsulated Czochralski
 (LEC) 247, 504
 – pulling rate 234
 liquid polymer 1233
 liquid silicon 259
 – impurity solubility 264
 liquid-crystal
 – columnar 923
 – nematic 923

- smectic 920
- twist grain-boundary phase 923
- liquid-crystal device 931
- liquid-crystal display (LCD) 590, 917
- addressing 933
- angle of view 933, 938
- bistable device 939
- deformed helix mode 947
- ECB-mode 932
- ferroelectric smectic display 939
- global market 1226
- in-plane switching (IPS) mode 937
- in-plane switching mode (IPS) 938
- materials 940
- multicomponent mixtures 942
- multiplexing 934
- optical properties 934
- passive matrix addressing 934
- relaxation time 934
- super-twisted nematic (STN) 935, 936
- thin-film-transistor twisted-nematic 937
- time response 934
- transmittance 934
- twisted nematic (TN) 935
- twisted vertically aligned (TVAN) 938
- vertically aligned nematic (VAN) 938
- viscous properties 934
- zenithal bistable device (ZBD) 939
- liquid-crystal material
 - fluorinated mesogens 948
- liquid-crystal mixtures 948
- two phase coexistence 942
- liquid-crystal phases
 - disc-like molecules 923
- liquid-encapsulated Czochralski technique 1349
- liquid-phase epitaxy (LPE) 9, 11, 271, 303, 328, 507, 873, 889, 1022, 1349
- second-generation CMT 308
- lithium ion batteries 225
- lithographic method 1002
- lithography 7
 - parallel 1233
- local chemical state
 - surface 375
- local density approximation (LDA) 757, 758
- local oxidation of silicon (LOCOS) 647
- localisation 13
- localised vibrational modes (LVM) 515, 517
- localized state 163, 174, 1109
 - potential barrier 180
- localized vibrational mode (LVM)
 - absorption 1349
- LOCOS isolation oxide 647
- logic circuitry
 - molecular-scale 1234
- logic device dimensions 630
- logic gates 1151
- log-pile structure 1007
- London equations 1197
- London penetration depth 1197
- London theory 1200
- long linear arrays 868
- long waveband (LW) 856
- longitudinal bias film 1345
- longitudinal gauge factor 1115
- longitudinal magnetic recording 1349
- longitudinal optical (LO) 759, 767, 1031
 - modes 741
 - phonons 55
- longitudinal recording
 - write head 1176
- longitudinal resistivity 43
- longitudinal TFT 1117
- longitudinal transition 1170
- long-range disorder 1349
- long-range orientational order 925
- long-wavelength detectors 876
- long-wavelength LO phonon 748
- loophole device 875
- loose-spin model 94
- LO-phonon coupling 784
- Lorentz oscillators 51
- loss coefficient 55
- low media noise 1170
- low remanence state 1187
- low- κ application 652
- low-dimensional structure 1349
- low-dimensional transport phenomena 1150
- low-energy ion scattering (LEIS) 346
- low-energy sputtering 671
- lowest unoccupied molecular orbital (LUMO) 1226
- low-field electron drift mobility
 - GaN 814
- low-frequency dispersion 209
- low-frequency noise 1349
 - characteristics 430
 - measurement 433
 - spectroscopy 430
- low-impedance devices 433
- low-loss dielectric material 998
- low-mobility materials 144
- low-pressure chemical vapour deposition (LPCVD) 481
- low-temperature co-fired ceramic (LTCC) 1270, 1349
- low-temperature luminescence 521
- low-temperature MBE (LTMBE)
 - GaAs 510
- low-temperature solution growth 1349
- low-temperature solution growth technique 238
- low-temperature-grown GaN 775
- LPE 1349
 - background impurities 272
 - characteristics 272
 - GaAs 276
 - growth modes 273
 - history 272
 - n-on-p structure 280
 - optoelectronic devices 276
 - p-on-n structure 280
 - second-generation CMT 308
- LPE growth 329
- LPE growth method
 - dipping system 274
 - graphite sliding-boat 274
 - tipping furnace 273
- LPE growth techniques
 - comparison 310
- LPE material
 - point defects 273
- L-pit 104, 105, 1349
- LSDA (local spin density approximation) 97
- LTCC 1279
- LTMBE GaAs
 - point defects 517
 - ultrafast optodetectors 510
- lucky drift (LD) model 39
- lucky electron 38
- luminescence 983, 1349
 - band-to-band 987
 - centres 985
 - deep-level 987
 - electronic configuration 985
 - excitation/emission cycle 988
 - optically stimulated 990
 - rare-earth 985
- luminescent ions 987

luminescent material 983, 992, 995
 LUMO – lowest unoccupied
 molecular orbital 1350
 Luttinger liquid (LL) state 1151
 LW 1350
 LW matrix arrays 868

M

Mach–Zehnder
 – interferometer 911
 – modulator 1076
 macrocyclic compound 1241, 1245
 macrocyclic molecule 1246
 macrodefects 315
 macromolecule
 – cylindrical graphitic 1148
 magnetic anisotropy 88
 magnetic annealing 1157, 1350
 magnetic coupling 1165
 magnetic field
 – effective 1187
 magnetic field applied continuous CZ
 (MCCZ) 267
 magnetic field applied CZ (MCZ)
 266
 magnetic film 1155, 1157
 – engineering 94
 – permanent 1166
 – switching time 1162
 magnetic free enthalpy 1187
 magnetic grains 1169
 magnetic layers
 – amorphous metallic 99
 magnetic material 1155
 – antiferromagnetism 90
 – ferrimagnetism 90
 – ferromagnetism 90
 – multi-domain structure 86
 – thick film 613
 magnetic permeability 48
 magnetic properties 79
 – thin film 660
 magnetic random-access memory
 (MRAM) 1156, 1185, 1350
 magnetic read head 92
 magnetic recording 1156
 – head 1158
 – media 92
 – microstructure 1169
 – read head 1188
 – surface 1169
 – transition 1182
 magnetic sector instrument 377
 magnetic semiconductors 883
 magnetic spacing 1350

magnetic splitting 99
 magnetic substance
 – fundamental property 81
 magnetic surface charges 81
 magnetic thin film
 – coupling strength 93
 magnetic transition 1350
 magnetic tunnel junction (MTJ)
 1156, 1185
 magnetic tunneling junction (MTJ)
 1350
 magnetic yoke 1159
 magnetism
 – classical 81
 – localized 81
 – transition metal alloys 97
 magnetization 1187
 – curve 1178
 – equation of motion 1187
 magnetoelastic energy 1157
 magnetoelectronics 99
 magneto-optical effects 744
 magnetoresistance 1350
 magnetoresistance (MR) 446
 magnetoresistance ratio 1186
 magnetoresistive sensors 730
 magnetoresistivity (MR) 1350
 magnetostriction 1157
 – constant 1160
 – nickel–iron alloys 1157
 magnetron sputtering 676
 magnification 351
 majority carriers 1350
 Marconi company 940
 mask processing 1003
 masking layer 1144
 mass flow controllers (MFC) 295
 master curve technique 204
 material 3
 – carbon-based 1220
 – charge transport 161
 – composition 167
 – disordered 161
 – GaN 14
 – Group III-nitride 12
 – homogeneous nonlinear 1077
 – magnetic 1155
 – material 1078
 – nonlinear 1084
 – organic 1084
 – organic thin film 708
 – photovoltaic 1098
 – resonant response 1079
 – second-order 1088
 – semiconducting 3
 – sensor-active 1241
 – thermal conductivity 391
 – thick film 721
 material characterization 139
 – sample preparation 362
 – X-ray diffraction 348
 material flow 109
 material parameter 808, 809
 – quadratic dependence 736
 – ternary 736
 material photosensitive 140
 material porous 1009
 material properties
 – Group III–V nitride semiconductors
 822
 material quality 537
 material systems
 – viscosity 401
 materials growth
 – diamond 242
 – Gallium Arsenide (GaAs) 242
 – Group III–V 242
 – Group II–VI 245
 – Group IV 240
 – Group-III nitrides 244
 – Halides 250
 – Indium and Gallium Antimonides
 244
 – Indium and Gallium Phosphides
 244
 – oxides 249
 – oxides/halides/phosphates/borates
 249
 – phosphates and borates 251
 – Silicon 240
 – Silicon Carbide 241
 materials LPE grown
 – arsenic-based materials 276
 – atomically flat surfaces 280
 – doping 277
 – Group III–V 276
 – Group II–VI 278
 – group IV 275
 – Group-III nitrides 277
 – Group-III–V antimonides 276
 – low-bandgap material 277
 – mercury cadmium telluride 278
 – mixed alloy layers 277
 – narrowgap Group II–IV compounds
 280
 – phosphorus-based materials 276
 – silicon 275
 – silicon carbide 275
 – silicon/germanium 275
 – widegap compounds 278
 materials MBE grown
 – GaAs/AlGaAs 296

- GaAs-based electronic devices 299
- Group-III nitrides 298
- Group-III–V nitrides 298
- InGaAs 297
- precursor 299
- quality 298
- quantum dot 298
- sapphire substrates 298
- Stranski–Krastanov mode 298
- materials MOCVD grown
 - $\text{Al}_{1-x}\text{Ga}_x\text{As}$ 286
 - antimonides 286
 - arsenides 286
 - GaN 287
 - Group II oxides 289
 - Group II–VI semiconductors 288
 - HgCdTe 288
 - nitrides 287
 - phosphides 286
 - selenides 289
 - sulfides 289
 - ZnSe 289
- materials science 4
- Matthiessen's rule 23, 696, 750, 766
- Mauguin condition 935, 936
- maximum efficiency 1098
- Maxwell–Boltzmann distribution 29, 663
- MBE 332
 - AlAs 292
 - commercial systems 295
 - GaAs 292
 - gas sources 295
 - Group III–V compounds 292
 - growth chamber 293
 - growth rate 291
 - growth systems 293
 - history 290
 - reaction mechanisms 291
 - $\text{Si}_{1-x}\text{Ge}_x$ film 488
- MBE growth
 - defects 294
 - substrate-preparation 294
- MBE technology
 - CMT 317
- MBE-grown CMT
 - alternatives 320
 - applications 320
 - device-quality 319
 - surface morphology 318
- MCZ silicon 266
- mean free path 392, 696
 - bulk electron 696
- mean grain size 1170
- mean phonon velocity 392
- mean-field approach 197
- mean-square relative displacement (MSRD) 1141
- mechanical properties
 - nitrides 757
 - thin film 660
- mechanical sensors
 - piezoresistive 728
 - thick film 728
- media flux 1169, 1350
- media noise 1350
- medium waveband (MW) 856
- medium-energy electron diffraction (MEED) 360
- medium-wavelength infrared (MWIR) 312
- melt-grown GaAs 507
- membrane
 - conductivity 1261
 - sensing 1245
 - solid state ionic 214
 - spin coating 1245
- memory architecture 1185
- memory chips 102
- memory device 1235
 - tunneling oxides 629
- memory storage technology
 - compact disk 651
 - controlled phase change 651
- memory technology 647
- mercury cadmium telluride (MCT) 278
- mercury-sensitized photo-CVD 1350
- MESFET (metal–semiconductor field-effect transistor) 549
 - transconductance 553
- mesogen 921, 1350
 - chemical constitution 944
- metal
 - bulk nonlinearity 1071
 - thermal expansion coefficient 398
- metal alloys
 - sheet resistivity 721
- metal gate electrode 634, 644
- metal oxide semiconductor (CMOS) 1229
- metal oxide semiconductor field effect transistor (MOSFET) 7
- metal semiconductor field-effect transistor (MESFET) 1350
- metal-free phthalocyanine 1254
- metal-induced gap states (MIGS) 155
- metal–insulator transition 1193, 1203
- metal–insulator–metal (MIM) 647
- metal–insulator–semiconductor (MIS) 704, 858
- heavily doped polycrystalline Si 633
- structure 633
- metallic film 696, 700
 - electrical conductivity 696
- metallic glasses 396
- metallic impurity 103
- metallic nanocomposites 1088, 1350
- metallic superconductor 1193, 1201
 - isotope effect 1198
- metallization 544
- metalloporphyrin
 - glass 1249
 - Langmuir–Blodgett film 1249
 - polymer 1249
- metallurgical-grade silicon (MG-Si) 257
- metalorganic chemical vapor deposition (MOCVD) 9, 10, 12, 271, 280, 325, 328, 614, 680, 754, 1022
 - epitaxy 1082
- metalorganic deposition (MOD) 614
- metalorganic molecular beam epitaxy (MOMBE) 295, 325, 328, 332, 511, 754
- metalorganic vapor phase epitaxy (MOVPE) 9, 280, 303, 328, 508, 874, 1102
 - CMT 312
 - heterodiode technology 881
 - reactor cells 284
- metal–oxide–semiconductor (MOS) 158, 696
- metals 25
 - electrical conductivity 396
 - electrical properties 25
 - resistivity 25, 411
 - sheet resistivity 721
- metal–semiconductor barrier 151
- metal–semiconductor contact 147, 148
 - depletion layer 150
 - J – V characteristic 416
 - patches 150, 151
 - photoelectric response 152
- metal–semiconductor field effect transistor (MESFET) 9, 513, 525

- metal–semiconductor interface
 - electrical properties 544
- metamorphic interface 157
- metastability 1350
- Metglass 91
- MgB₂
 - critical temperatures 1214
- micro screen 718
- microcrystalline (μc) 1108
- microcrystalline silicon
 - hydrogenated 581
- microcrystals 1086
- microdiffraction 358
- micro-electromechanical system (MEMS) 226, 537, 546, 557
 - diamond 551
 - thin film 613
- microelectronic applications
 - thermally grown SiO₂ 636
- microelectronic circuits 717
 - Si-based 625
- microelectronic devices
 - dielectric materials 625
 - scaling 640
- microelectronic materials 636
- microelectronics 80, 95, 99
- microhardness 739
- microindentation methods 758
- micromachined Si 1007
- micromachining 1244
 - technique 1004
- microphotonics 1350
- micropipes 543
- microprocessor chip 1229
- microprocessor performance 628
- microprocessor unit (MPU) 1229
- microprocessors
 - scaling 626
- micro-pulling down technique 249
- microscopy
 - optics 351
- microstrip design 1279
- microstructure glass 1257
 - sensor 1257
- micro-via 1278
- microwave circuit 1271
- microwave device 90, 276
- microwave transistors 525
- mid-gap-states 1068
- migration energy 214, 219
- Miller's rule 1066, 1351
- milling 611
- miniaturization 214
- miniband 1026, 1351
 - break 1029
 - injector region 1034
- minority carrier 912, 1351
 - drift length 861
 - lifetime 318
 - silicon-based electronic device 454
- MIPS (millions of instructions per second) 1236
- MIRIADS (miniature infrared imaging applications development system) 882
- mirror 692
- MIS 1351
- MIS diode 640
- misalignment angle 1210
- misfit 11
- misfit dislocation 356, 511, 1026, 1351
 - transition metal impurity decoration 369
- mixed conducting 224
- mixed conductor 213
 - electrical properties 218
- mixed crystal
 - charge carrier 168
 - compositional fluctuation 168
- mixed state 1351
- mixture rule 33
- MMICs (monolithic microwave integrated circuits) 514
- Mn-doped AlN 786
- mobile carrier
 - concentration 216
 - sublattice 217
- mobile ion
 - electrical properties 218
- mobile monitoring 1243
- mobility 19, 750, 870, 1351
 - degradation 632
 - degradation coefficients 427
 - edge 163, 175, 179, 566, 1111, 1351
 - effective 1112, 1113
 - field-effect 1111
 - gap 163, 173, 1109, 1110, 1351
 - of electrons 177
- mobility–lifetime product 1128
- MOCVD 286, 1351
 - basic reaction kinetics 281
 - characteristics 281
 - precursors 283
 - technique 331
- mode
 - AlAs-like 741
 - GaAs-like 741
- mode hops 905
- model-dielectric-function (MDF) 750
- modulated photoconductivity (MPC) 142
- modulation doped field-effect transistors (MODFETs) 795
- modulation doping 10
- modulation transfer function (MTF) 866, 1121, 1123, 1131, 1351
 - aperture 1134
 - detector material 1134
- modulation-doped FET (MODFET) 297
- modulation-doped GaAs/AlGaAs 297
- modulation-doped heterointerfaces 754
- modulator 1036
 - electroabsorption 907, 1037
 - electro-optic effect 911
 - InGaAsP-based 910
 - linewidth enhancement 911
 - optical 888
 - polarisation insensitive 909
 - refractive index 911
 - waveguide configuration 909
 - wave-guide structure 910
 - Y-junction 911
- modulator device 1029
- molecular
 - chirality 921
 - density 737
 - dipole relaxation time 200
 - dispersion 960
 - dynamics (MD) 568
 - manufacturing 1219
 - memory 1219
 - switch 1219
 - wires 924
- molecular beam epitaxy (MBE) 9, 10, 12, 152, 271, 290, 317, 325, 328, 481, 661, 683, 754, 889, 1022, 1082
 - GaAs 509
- molecular electronics 1219, 1351
 - architecture 1233
 - electroactive compound 1224
 - manipulating material 1233
- molecular recognition 1231
- molecular reorientation 1351
- molecular-beam epitaxy (MBE) 1351
- molecularly doped polymer (MDP) 162, 180, 958

molecularly doped polymers (MDP) 1351
 molecular-scale technology 1230
 momentum-flipping collisions 392
 monolayer (ML) 291, 1351
 monolithic solder 1278
 monomolecular recombination 143
 monopolar conduction 36
 monosilane (SiH_4) 257
 – production 258
 – pyrolysis 258
 Monte Carlo simulations 805
 – electron transport 813
 Moore's Law 1351
 Moore's law 7, 8, 626, 1229
 morphological characteristics 659, 690
 MOS 1351
 MOS transistor 1185
 Moseley's law 1351
 MOSFET
 (metal–oxide–semiconductor field-effect transistor) 410
 – injection efficiency 869
 – SiC 547
 Mott's law 172
 MOVPE 1351
 MPCVD technique 1352
 MQW lasers 297
 MRAM technology 1155
 MTF model 1136
 multi-component material
 – sputtering rate 376
 multicore wire 1196
 multicrystalline silicon
 – loss of efficiency 1099
 – module efficiency 1099
 multi-crystalline silicon (mc-Si) 1352
 multidomain structure
 – magnetic material 86
 multijunction cell 1098, 1100
 multilayer ceramic (MLC) 611
 multilayer device 968
 multilayer film 578
 – quantum size effects 571
 multilayer heterojunction (MLHJ) 319
 multilayer media 1181
 multilayer structure 1235
 multioscillator model 748
 multiple quantum well (MQW) 69, 124, 332, 528, 857, 1024, 1352
 – laser 530
 – nonlinear response 1081
 – semiconductor MQW 1082

multiple reflection approximation 473
 multiplets 985
 multiplexer 858
 – design 879
 multistrain well 910
 multiwall nanotubes (MWNT) 1148
 – ballistic conductance 1150
 – electrical transport 1150
 – inter-layer coupling 1150
 MW 1352
 MW detectors 868

N

N doping 847
 nanocomposites
 – metallic 1088
 nanocrystal 1087
 – inorganic semiconductor 1087
 – nonlinear susceptibility 1088
 nanocrystalline (nc) 1108
 nanoelectronic circuit 1233
 nanoelectronic device 708, 1151
 nanoengineering 997
 nanofibrous membrane 1259
 nanoflash device 1235
 nanoindentation methods 758
 nanolithography (DPN)
 – dip-pen 1233
 nanomaterials 1244
 nanoparticle 1235
 nanoparticle system 1070
 nanoplotter 1233
 nanoporous array
 – high-aspect 1005
 nanostructure 661, 692, 997
 – oxide 1152
 – semiconductor 1150
 nanotechnology 565, 1147
 nanotube
 – armchair 1225
 – carrier mobility 1148
 – electrical conductivity 1148
 – electronic structure 1150
 – electrooptics 1152
 – logic gates 1151
 – semiconducting 1151
 – solar cell 1152
 – thermal conductivity 1148
 – zigzag 1225
 narrow bandgap 1352
 narrow gap III–V compounds 882
 NASICON 224
 native defect 125, 767, 1352
 native interstitials 130
 nearest-neighbor hopping (NNH) 170
 near-field technique 1144
 negative differential resistance (NDR) 524, 776, 1352
 negative electron affinity (NEA) 556
 negative temperature coefficient (NTC) resistance 729
 negative-bias temperature instability (NBTI) 640
 nematic
 – birefringence 943
 – dielectric anisotropy 943
 – phase 919, 1352
 – viscosity coefficients 929
 nephelauxetic effect 68
 Nernst equation 223
 net electric dipole moment 190
 NETD 1352
 neutral-impurity scattering 453
 neutrality condition 444
 neutron diffraction 216
 neutron transmutation doping (NTD) 260
 n-FET drive current 628
 Nichia Chemical Industries 12
 nitride
 – alloys 791
 – epitaxy 754
 – material system 753
 – semiconductor 805
 – semiconductors 754
 nitride-based
 – compound 1023
 – devices 754
 – semiconductor lattice parameters 756
 – technology 366
 nitrides
 – mechanical properties 757
 – optical properties 777
 – steady-state electron transport 775
 – thermal properties 761
 nitriding 678
 n-MOS 7
 n-MOSFET 10
 NO_2 sensor 1250
 noise equivalent temperature difference (NETD) 878
 noise sources
 – generation–recombination 862
 – Johnson–Nyquist 862
 – modulation 862

- Shockley–Read 863
 - surface trapping 862
 - noise transfer efficiency 1133
 - noise voltage 1170
 - nonabsorbing films 696
 - noncrystalline material 705
 - nondegenerate material 707
 - nonepitaxial film 683
 - nonlinear directional coupler 1352
 - nonlinear Fabry–Perot–interferometer 1352
 - nonlinear figure of merit 1352
 - nonlinear
 - Kramers–Kronig–transformations 1352
 - nonlinear Mach–Zehnder modulator 1352
 - nonlinear material 1084
 - optical 1089
 - quantum-confined inorganic semiconductor 1087
 - refractive index 1076
 - nonlinear material figures of merit 1352
 - nonlinear optical medium 1352
 - nonlinear optical switching 1352
 - nonlinear periodic structure 1352
 - nonlinear phase shift 1352
 - nonlinear polarization 1068, 1353
 - nonlinear refraction 1079
 - nonmagnetic metals 23
 - non-oxide compounds 674
 - nonradiative
 - decay 1028
 - process 984
 - recombination 567
 - nonreversing heat flow (NHF) 404
 - nonstoichiometric oxides 1202
 - non-stoichiometry 1353
 - nonthermal energy exploration telescope 839
 - nonvolatile information
 - thin film 613
 - nonvolatile memory devices
 - dielectrics 649
 - Nordheim coefficients 27
 - Nordheim’s rule 26
 - normal–insulator–superconductor (N-I-S) 1199
 - current–voltage characteristic 1199
 - n^+ -p diodes 871
 - n-Si
 - extinction coefficient 474
 - refractive index 474
 - n-type carrier concentration 792
 - n-type channel MOS 539
 - n-type conductivity 1353
 - n-type diamond
 - electron densities 552
 - nitrogen-doped 552
 - pn-junction 552
 - n-type GaAs 504, 513
 - ohmic contact 524
 - n-type GaN 767
 - n-type HgCdTe 864
 - n-type semiconductor 767
 - n-type silicon
 - electron mobility 456
 - neutron transmutation doping 260
 - phosphine (PH_3) 260
 - n-type ZnO 849
 - nuclear magnetic resonance (NMR) 568
 - nucleation
 - activation energy 402
 - density 546
 - field 1180
 - temperature 111
 - Nucleation model 587
 - nucleation temperature 1353
 - numerical aperture 1353
 - Nyquist frequency 1128, 1135
 - Néel temperature 1203
 - Néel-Arrhenius formula 86
-
- O
-
- octahedral ferroelectric 605
 - off-chip interconnect 1278
 - off-track read profile 1177
 - Ohm mobility 456
 - OLED blended single layer
 - quantum efficiency 1227
 - OLED display
 - global market 1226
 - Onsager model 1353
 - Onsager relations 931
 - opal PBG structure 1008
 - opaline lattice 1009
 - optical absorption 141, 576, 789
 - coefficient 1353
 - edge 576
 - sub-bandgap 141
 - optical amplification
 - gain 906
 - optical axis 920
 - optical bistability 1076, 1353
 - optical communications 914
 - optical constant
 - Drude approximation 476
 - model 476
 - optical constants 1353
 - optical device
 - figures-of-merit 1068
 - nonlinear 1076
 - optical dielectric constant 748
 - optical disk 1032
 - optical effect
 - thin films 692
 - optical emission spectroscopy (OES) 582, 1353
 - glow discharge 376
 - optical feedback 897
 - optical fiber 12
 - attenuation 1068
 - carrying capacity 1075
 - material 1068
 - sensor 1257
 - transmission rate 1075
 - optical gap 60, 570, 571, 577, 1114, 1353
 - optical glass 50, 74
 - optical integrated circuit 1068, 1069
 - optical limiter 1353
 - optical lithography 1152
 - optical material 74, 75
 - silicon 472
 - optical media 1143
 - optical memory devices 565
 - optical mode 740, 748
 - optical network 1075
 - optical nonlinearity 1067
 - optical phonon scattering 28
 - optical phonon scattering rates
 - GaAs 814
 - GaN 814
 - optical poling 1069
 - optical properties 47, 565, 661, 692, 1353
 - aluminium nitride 786
 - a-Si:H 571
 - bound excitons in GaN 782
 - chalcogenide glasses 572
 - defect-related transitions in GaN 785
 - gallium nitride 778
 - GaN 785
 - indium nitride 789
 - nitrides 777
 - SiGe 488
 - thin films 692
 - optical pumping 890, 984

optical sensor 1251, 1256
 optical signal processing
 – broadband 1076
 – wavelength-selective nonlinear 1076
 optical storage 12
 optical switching 1078
 – phase shift 1079
 optical system integration 1002
 optical texture 920, 1353
 optical thickness 694
 optical transition
 – emission 1028
 – interband 1026
 – matrix 60
 – transition rate 1026
 optical transparency 545
 optically detected magnetic resonance 1353
 optically stimulated luminescence (OSL) 990
 optical-phonon scattering 771
 optoelectronic applications 766
 optoelectronic device 411, 735
 – detector 911
 – semiconductor 911
 – strain in 894
 optoelectronic diamond 551
 order parameter 924, 948, 1353
 order–disorder ferroelectrics 198
 order–disorder model
 – ferroelectric behaviour 600
 order–disorder transition 198
 organic compound
 – macroscopic properties 1219
 organic device
 – nanoscale 1234
 organic film 1230
 organic glasses 162, 180
 organic laminant 1270
 organic layer patterning 1232
 organic light emitting display 1219
 organic light-emitting diode (OLED) 1107, 1226
 organic material 1219, 1241
 – nonlinear response 1084
 organic photoconductor (OPC) 1353
 – architecture 967
 – function 967
 – material 967
 organic photoreceptor 953, 974, 1353
 – charge-transport layer (CTL) 974

organic semiconductor 1226
 – carrier concentration 1113
 – zero-field mobility 181
 organic solvent sensitivity 1254
 organic substrate module circuit 1279
 organic thin film material 708
 organic thin layer
 – dielectric constant 1253
 organometallic vapor-phase epitaxy (OMVPE) 878
 organo-silicate glasses (OSG) 653
 organosilicon 1231
 orientation preferred 361
 orientational order 946
 oscillator strength 1353
 out-of-plane coherence length 1207
 overcoat
 – diamond-like carbon 976
 – layer 975
 – polyurethane 976
 – silsesquioxane 976
 overlap zone-melting 838
 oxidation-induced stacking fault (OSF) 103, 1353
 oxide capacitance 422
 oxide film 674
 oxide glass 52
 – empirical rules 390
 oxide nanostructure 1152
 oxide trap density 432
 oxide/nitride (ON)
 – dielectric stacks 649
 oxide/nitride/oxide (ONO)
 – dielectric stacks 649
 oxygen ion conductivity 220
 oxygen precipitates 1353
 oxygen sensor 223, 1243
 oxygen stacking fault (OSF) 101
 oxygen vacancy 850
 – concentration 215
 oxygen-doped AlN 786
 oxygen-related defects 759
 ozone monitoring 1242

P

package style 1270
 packaging 727
 packaging system
 – conjugate 1268
 paraelectric 1353
 paraelectric phase 599, 605
 parallel lithography 1233
 paramagnetism 96

parasitic
 – capacitance 628, 633
 – resistance 633
 parity selection rules 66
 partial response maximum likelihood (PRML) 1170
 – recording channel 1353
 particle flux 1025
 particle–material interaction 345
 passivation 1353
 passivation layer
 – thick film 723
 passive electronic components
 – thick film technology 724
 passive matrix addressing 1354
 Pauli principle 1199
 Pb-free
 – soft error 1278
 – solder 1278
 PC fabrication
 – radiation patterning 1002
 p-doping 13
 PECVD technique 1354
 Peierls distortion 1150
 Peltier cooling 1130
 penetration depth (λ_L) 1354
 penetration depths 1207
 percolation 123
 – distance 176
 – parameter 183
 perfect diamagnetism (Meissner effect) 1354
 performance enhancing electronics 538
 perimeter-bonded die 1270
 periodic structures 1076
 peripheral package 1268
 permalloy 1159, 1354
 permanent magnetic film 1166
 permanent magnets 92
 permeability 88, 1161
 permeation rates 225
 permittivity 630, 1077, 1354
 – of free space 188
 – relative 48, 188
 – vacuum 48
 perovskite (CaTiO_3) 601
 perpendicular
 – alloys 1178
 – anisotropy 1180
 – critical magnetic field 1178
 – geometry 1175
 – magnetic recording 1354
 – magnetization 1179
 – media 1178
 – recording 1178

- spin valve head 1176
- write head 1175
- perturbation Hamiltonian
 - time-dependent 986
- PES technique (photoemission spectroscopy) 1354
- p-FET drive current 628
- phase diagram
 - superconductor 1203
- phase equilibria 304
- phase sensitive detection 991
- phase transformations 402
- phase-change
 - band gap 1140
 - material 1140
 - recording 1140
- phase-change data storage 1139
- phonon 47
 - concentration 386, 387
 - coupling 782
 - dispersion 759
 - longitudinal optical (LO) 1031
- phonon scattering
 - non-polar 28
 - polar 28
 - temperature dependence 29
- phonon Umklapp scattering 762
- phonon–phonon anharmonic interaction 392
- phonon–phonon scattering 392, 762
- phonons 1354
- phosphor 983
 - luminescence efficiency 994
 - optical display 994
 - persistent 993
 - strip 993
 - X-ray storage 993
- phosphorescence 984, 1354
- phosphorus-based materials 276
- phosphorus-doped diamond films 540
- photo-catalysis 1354
- photochromic 1354
 - effect 1071
- photoconduction threshold 964
- photoconductive (PC) 278
- photoconductive array 860
- photoconductive detectors 860
 - Johnson-noise limited 863
 - long wavelength 866
 - n-type 863
 - sweep-out condition 863
- photoconductive gain 860
- photoconductivity 4, 137, 576, 972, 1354
 - experiment 138
 - layer 969
 - material 308
 - material properties 1123
 - photoreceptor 972
 - recombination rate 139
 - semiconductors 176
 - transient 142
- photoconductor 139, 953, 1123, 1135
 - amorphous 1124
 - band gap 1123
 - CdZnTe 1128
 - charge transport properties 1131
 - electronic noise 1134
 - energy absorption coefficient 1132
 - imaging properties 1130
 - linear attenuation coefficient 1131
 - long-wavelength array 861
 - loop 976
 - material properties 1130
 - PbI₂ 1129
 - poly-PbO 1129
 - quantum efficiency 139
 - responsivities 862
 - tribocharged 954
 - X-ray absorption profile 1132
- photocurrent 1097, 1354
- photodarkening 1072, 1354
- photodetector 735, 750, 888
 - active region 913
 - avalanche 913
 - band structure 912
 - near infrared 1034
 - optical 912
 - optimum band gap 912
 - photoconductive 912, 913
 - photocurrent 913
 - quantum-well inter-sub-band 1035
 - tunnelling 912
 - visible 1034
- photodiode
 - background-limited 869
 - dynamic resistance 869
 - transconductance 869
 - unipolar avalanche 1035
- photodiode junctions 874
- photodischarge 965
 - rate 966
- photoelectric effect 344
- photoelectrons
 - energy distribution 153
- photoexcited carriers 318
- photoexcited transport molecule 960
- photogeneration
 - coating 963
- photoinduced
 - absorption (PA) 567
 - anisotropy 1071
 - charge generation 956
 - degradation 591
- photoinduced degradation 1354
- photoinduced discharge characteristic (PIDC) 1354
- photoinduced effects 1355
- photoisomerization 1085, 1355
- photolithography 1003, 1232
 - serial patterning 1003
- photoluminescence 1355
- photoluminescence (PL) 521, 779, 782, 837, 847, 984, 1114
 - amorphous semiconductors 570
 - detector 991
 - long term 991
 - measuring system 991
 - short term 991
- photon 1355
- photon density 897
- Photon Factory 1142
- photon recycling technique 992
- photonic band gap (PBG) 997, 998
- photonic crystal (PC) 997
 - alumina 1005
 - architecture 999
 - cell structure 1000
 - design limitation 1001
 - device 997
 - electrochemical etching 1004
 - fabrication method 1005
 - fabrication methodology 1001
 - lithographic fabrication 1007
 - lithography fabrication 1002
 - magnetic 1013
 - materials and criteria 1000
 - nanometer-scale 1001
 - optical functionality 1009
 - optical response 1011
 - patterning 1003
 - PBG structure 1001
 - physical architecture 1000
 - physical structure 1012
 - self-ordered porous 1005
 - semiconductor-based 1011, 1012
 - structure 998, 1001, 1005
 - structure 3-D PBG 1007
 - tunable 1011
 - tuning 1012

- photonic crystal (PC) material
 - optical property 1000
- photonic crystal fabrication
 - charged particle 1003
 - ion-beam technique 1005
 - mask processing 1003
 - self-assembly methods 1008
- photonic device 997
- photonic glass 1355
- photonic glasses
 - optical nonlinearity 1063
- photoreceptor 965, 970, 972, 976
 - amorphous semiconductor 956
 - architecture 967
 - back-coat polymer 975
 - characterization 965
 - charge generation 957
 - charge transport 957, 967
 - charge transport layer 975
 - charge-generation layer 970
 - chemical resistance 977
 - conductive layer 969, 976
 - dark conductivity 956
 - dark decay 965
 - dark injection 969
 - device architecture 953
 - discharge 966
 - drift mobility 959
 - dual layer 957
 - electrical characteristic 953
 - electrical uniformity 977
 - electron transport 968
 - electrophotographic 967
 - fabrication 976
 - hopper coating 968
 - inorganic material 956
 - layer 967
 - mechanical strength 978
 - organic 953
 - overcoat layer 975
 - photoconductivity 972
 - photodischarge rate 978
 - photosensitivity 978
 - quantum efficiency 957
 - spatial uniformity 977
 - surface charge 956
 - xerographic 969
- photorefractance 778
- photorefractive beam fanning 1355
- photosensitive material 140
- photosensitivity 966, 1355
- photosensors
 - screen printed 730
 - thick film sensors 730
- photothermal deflection spectroscopy (PDS) 567
- photothermal ionization spectroscopy (PTIS) 519, 1354
- photovoltaic (PV) 278
 - conversion efficiency 1228
 - organic compound 1228
- photovoltaic array
 - HgCdTe 858
- photovoltaic cell
 - band diagram 1096
 - diode characteristic 1096
 - efficiency 1096–1098
 - figures of merit 1096
 - maximum power 1097
- photovoltaic detector 868
- photovoltaic device 731, 1095, 1355
 - current-voltage characteristics 868
- photovoltaic material 308
 - multigrain silicon 1098
 - single crystal 1098
- photovoltaic module 1355
- photovoltaic solar cell 1095, 1355
 - large-scale implementation 1105
- photovoltaic structure
 - alloy composition 1102
 - device design 1102
- phthalocyanine 1250
 - polymorphic 1248
 - rare-earth metal 1249
- phthalocyanine film 703
- phthalocyanine thin film 1254
- physical transformation 400
- physical vapor deposition (PVD) 330, 661, 1127
- physical vapor transport (PVT) 326, 542
- piezoelectric
 - actuator 617, 1014
 - ceramic 603
 - coefficient 604
 - device 617
 - effect 1013
 - material 617
 - scattering 767
 - sensor 617
- piezoelectric acoustic-phonon scattering 771
- piezoelectric elastic properties 618
- piezoelectric polarization 1355
- piezoelectricity 1355
- piezoresistance 1113, 1114
- piezoresistive sensors 729
- piezoresistivity 1355
- pigment material 963
- pinned film 1164, 1355
- pinning
 - centres 1212
 - effect 1163
 - efficiency 1212
- Piper–Polich method 246, 334
- pitch helix 921
- pixel rate 865
- pixel size 955
- pixie dust 1355
- planar diode 876
 - (I–V) characteristics 869
- planar lightwave circuit (PLC) 998, 1355
- planar process 7
- Planck's law 892
- planetary reactor 285
- plasma cleaning 364
- plasma display panel (PDP) 590, 994
- plasma etching 1004
- plasma frequency 55, 477
- plasma-assisted MBE (PAMBE) 754
- plasma-enhanced chemical vapor deposition (PECVD) 581, 679, 1108
- plasma-enhanced milling 874
- plasmon 1355
- plasmon energy 154
- plastic deformation 1355
- plastic electronics 1355
 - semiconducting 1225
- plastic encapsulation 1281
- plastic transistor 1219
- plating impurity 1272
- platinum resistance thermometers (PRTs) 729
- platinum thick film 722
- p-MOS 7
- p–n diode structure
 - I–V characteristic 426
- p–n junction 124, 898
 - reverse-biased 899
- pocket calculators 591
- point defect 1355
 - electrically active 434
 - surface 125
 - thermophysical properties 104
- point defect concentration 109
- point defect configuration 103
- point group 920
- point source 664

- Poisson's equation 1355
Poisson's ratio 1356
– semiconductors 740
polar crystal 598
polar dielectrics 598
polarity 1356
polarizability 48, 629
polarization 188, 929, 1063
– intensity-dependent 1078
– microscopic origins 190
– nonlinear 1077
polarization ratio 1356
polaron model 962
poles of a write head 1356
poling 612
poling mechanism 1069
pollutant gas
– detection 1242
pollution monitoring 1242
poly(methylphenylsilylene), PMPS
– glass transition temperature 961
poly(N-vinyl carbazole) PVK 972
polyacetylene 1222
polyaniline 1223
– conductivity 1256
polycrystalline
– alloy 396
– AlN 773
– conductivity σ 31
– CVD diamonds 540
– CZT 1128
– diamond 551
– dielectric 610
– film 31, 32, 687
– GaAs 506, 1101
– growth 280, 361
– interface states 430
– material 30, 1210
– mercuric iodide 1124
– piezoelectric 610
– pyroelectric 610
– seed crystal 503
– silicon 1114
– silicon emitter 432
– superconductivity 1210
– thin films 685, 1095
– vapour-deposited 686
polycrystalline solid 1356
polydiacetylene
– figures of merit 1086
– nonlinear properties 1085
polydimethylsiloxane (PDMS) 1233
polyelectrolyte 1231
polyethylene
– crystallinity 394
polygonization 1356
poly-HgI₂ 1132
polymer 1063, 1069
– backbone 1222
– blend composition 395
– conducting 1241, 1255
– conductivity 1222
– conjugated 180, 961, 1224
– effective thermal conductivity 395
– electrical conduction 1223
– glass-transformation kinetics 404
– heat capacity 389
– molecularly doped 958
– semiconductive 1224
– substrates thick film 720
– thick film conductors 724
– thick films 718, 723, 1356
– transistors 431
polymer conducting
– physical property 1241
polymeric
– composite thermal conductivity 395
– ferroelectric polyvinylidene fluoride 607
– ion-selective electrode 1257
– membrane 1257
– organic material 1221
– sensor 1257
polymorphous silicon 142
poly-PbI₂ 1129
polyphenylenevinylene (PPV) 1223, 1226
polypropylene structure 389
polypyrrole 1223
polysilicon
– emitter bipolar junction transistor (PE BJT) 410
– granular 258
– production 258
– randomly stacked 258
– seeding 259
polythiophene 1223
polyvinylcarbazole (PVK) 180
polyvinylidene fluoride (PVDF) 607, 729
Poole–Frenkel emission 703
porous array
– domain size 1005
– pore shape 1005
porous material 1009
porphyrin
– synthetic 1249
positive temperature coefficient (PTC) thermistors 730
positron annihilation 1356
positron emission tomography (PET) 994
post-growth heat treatment 304
potassium dihydrogen phosphate (KDP) 600
power
– devices 538
– dissipation 430
– electronics 555
– miniaturization 214
– sources 214
– stabilizer 1356
power added efficiency (PAE) 549
precision doping 261
precursor 283, 680, 1356
– diethyl telluride (DETe) 313
– diisopropyltelluride (DIPTe) 313
– di-methyl cadmium (DMCd) 314
– Group II 285
– Group V 285
preferred orientation 361
pressure sensors 729
primary bombarding particles 383
primary ion species 379
printed capacitors 725
printer
– liquid-toner-based 976
process integration 1356
process technology
– 0.13 μm 1229
profiling techniques 127
proper ferroelectric 1356
proton conductivity 220
pseudogap (cuprate superconductors) 999, 1203, 1204
pseudomorphic HEMT (pHEMT) 10, 526
pseudomorphic interface 157
pseudomorphic layer 1356
pseudopotential linear muffin-tin orbital (FP-LMTO) 757
p-type
– dopant 845, 848
– doping 849, 850
– GaAs 845
– HgCdTe 870
– silicon 260
– ZnBeSe 849

p-type conductivity 1356
 p-type diamond
 – electrical resistivity 551
 – nitrogen-doped 552
 – pn-junction 552
 p-type ZnO
 – complexes 850
 – ionization energy 850
 – resistivities 850
 pull rate 106
 pulse width 1356
 pulsed laser deposition (PLD)
 754
 pump–probe experiment 1081
 pure GaAs
 – properties 501
 purification 1149
 purified Si 444
 Pyrex glass 1069
 pyroelectric coefficient 604
 pyroelectric device
 – radiation detector 620
 pyroelectric effect 620, 1356
 pyrolytic boron nitride (pBN) 505,
 1356
 PZT ceramic 618

Q

q-DC behaviour 209
 quadratic Stark effect 1029
 quantum ballistic transport 42
 quantum cascade laser (QCL)
 889, 1021, 1033, 1356
 quantum communication devices
 80
 quantum computer 80, 98
 quantum conductance 1150
 quantum confined Stark effect
 (QCSE) 908, 1356
 quantum confinement 1021, 1083
 quantum Dot (QD) 1356
 quantum dot (QD) 298, 332, 565,
 889, 1037, 1087
 quantum dot laser 894
 quantum dot layer 904
 quantum efficiency 139, 868, 872,
 964, 1100, 1121, 1123, 1357
 – differential 900
 – internal 900
 quantum Hall effect (QHE)
 418
 quantum interference
 – Josephson junctions 1200
 – model 94
 quantum resistance 1150

quantum size
 – amorphous semiconducting
 multilayers 577
 quantum size effect 1023
 quantum well (QW) 332, 521, 522,
 565, 577, 754, 889, 894, 1021,
 1357
 – active region 899
 – alloy semiconductor 1028
 – application 1032
 – biased 1030
 – carrier confinement 898
 – conversion efficiency 899
 – density of states 894
 – electronic state 1024
 – GaAs/AlGaAs 1029
 – inter-sub-band photodetector
 (QWIP) 1035
 – laser 894
 – luminescence peak 1028
 – quaternary alloy 1032
 – strained-layer 1032
 – threshold gain 899
 quantum well infrared photodetectors
 (QWIP) 488
 quantum well inter-subband
 photodetector (QWIP) 1357
 quantum well laser 883
 quantum well structure 11
 quantum wire 41, 565, 894, 1037,
 1357
 quantum-confined Stark effect
 (QCSE) 1028, 1036, 1357
 quantum-dot glass 1087
 quantum-mechanical confinement
 633
 quantum-mechanical tunneling
 629
 quantum-size effect 1357
 quantum-well device 276
 quarter-wave stack 695
 quartz microbalance 1228
 quasi Fermi level 140, 893
 quasi particles 1357
 quasi-Fermi levels 1357
 quasi-particle band gap 154
 quasi-static measurements 421
 quaternary
 – alloy 750, 1022, 1032
 – binary parameters 736
 – composition parameters 735
 – Hall mobility 751
 – optical modes 741
 – unit cells 741
 qubit 80
 quench anneal (QA) 245, 305

R

radar 10
 radial distribution function (RDF)
 568
 radiation detector 620
 radiation patterning 1002
 radiation resistance 1357
 radiation–material interactions 344
 radiative process 984
 radiative recombination 567, 572
 Radio Corporation of America 940
 radio frequency (RF) 592, 1271
 radioactive Pb²¹⁰ 1277
 radio-frequency (RF) circuits 1357
 radio-frequency (RF) sputtering
 675
 radioluminescence 984
 Raman crystallinity 588
 Raman fiber amplifiers 1068
 Raman scattering
 – stimulated 1068
 Raman-active phonon modes 759
 randomly oriented particles 33
 randomly stacked polysilicon
 – Siemens method 258
 rapid thermal annealing (RTA) 514
 rare-earth (RE) ion 68
 rare-earth spectroscopy 991
 Rayleigh coefficient 86
 RC delay time response 627
 reactive evaporation 667
 reactive sputtering 581, 674, 1357
 reactor cell 1357
 reactor planetary 285
 read channel 1170
 read head 1357
 readout integrated circuit (ROIC)
 858
 receptor technology 958
 rechargeability 225
 reciprocal lattice 354, 1357
 recombination 146, 862, 1357
 – centers 140
 – nonradiative 900
 – process 176
 recombination radiative 567, 572
 recording 1139
 – density 1181
 – film decay rates 1173
 – film thickness 1159
 – layer 1144
 – media noise 1170
 – ultra-high 1181
 recording medium
 – magnetically saturated 1174

- recrystallization 306
 rectangular columnar phase 924
 rectifier 1219
 reduced dislocation density 506
 re-entrant phase 924
 reflectance 48, 694
 reflection coating 696
 reflection coefficient 70
 Reflection difference spectroscopy (RDS) 317
 reflection high energy electron diffraction (RHEED) 291, 344, 510
 reflectometer 1357
 reflow 1357
 reflow soldering 727
 refractive index 47, 52, 75, 694, 1077, 1086
 – average 52
 – complex frequency-dependent 189
 – intensity-dependent 1065, 1078
 – MQWs 1083
 – nonlinear 1077
 – nonlinear changes 1077
 – oxide glass 52
 – semi-empirical 1066
 – tuning 1011
 – ultrafast nonlinear response 1077
 refractory compound 213
 refractory layer 976
 relative dielectric permittivity 629
 relative permittivity 48, 188
 relative sensitivity 382
 relative susceptibility
 – frequency-dependent 191
 relaxation
 – ferroelectric 199
 – frequency 202, 205
 – rate 200
 – response 192, 194, 199
 – responses 196
 relaxation time 401
 – approximation 29, 448
 – dipole density fluctuation 200
 reliability of dielectrics 640
 remanent magnetization 1160, 1173, 1357
 remanent polarization 1357
 remote impurity scattering 40
 residual resistivity 23
 resistance
 – contact 415
 – junction 1219
 – limit 1196
 – metal-semiconductor 415
 – modulation 1113
 resistance temperature devices (RTDs) 310
 resistive evaporation 665
 resistive transition 1209
 resistivity 24, 32, 418, 1357
 – effective 23
 – index 25
 – longitudinal 43
 – measurement 451
 – mixture rule 34
 – residual 23
 – strain-induced 1114
 – temperature coefficient 24
 – Van der Pauw technique 413
 resistor
 – thick film 722
 resistor trimming 723, 726
 resolution 351, 1358
 resonance response 191
 resonant cavity LED (RCLED) 529, 906, 1033
 resonant nonlinear response 1358
 resonant sensors 729
 resonator
 – micro-machined 1014
 response function 1184
 responsivity 1358
 reststrahlen band 748
 retrograde solidus 1358
 RF circuit
 – flip chip 1277
 – interconnect 1277
 RF device
 – packaging material 1272
 RF magnetron sputtering 614
 RF packaging
 – substrate 1279
 – thermal issues 1284
 RF wirebond
 – digital 1273
 – materials 1273
 RHEED 1358
 Richardson
 – constant 149
 – formula 1030
 RIE (reactive-ion etching) 544
 ring coating 968
 Ritland plot 401
 Rochelle salt 598
 rod-like molecules 929
 ROIC 1358
 rotational viscosity 949
 rotational viscosity coefficient 931, 935
 rotaxane molecule 1234
 routing 1076
 R–T method 1358
 Rutherford backscattering (RBS)
 – spectrometry 343
 Rutherford back-scattering (RBS) 792
 Rutherford scattering 1358
-
- ## S
-
- SAM film 1247
 sapphire 13, 277, 298
 sapphire substrate 315, 760, 878
 saturable absorber 64
 saturated vapor pressure (SVP) 283, 1358
 saturation
 – flux density 1159
 – magnetization 82
 – parameter 64
 – polarisation 599
 saturation intensity 1358
 saturation magnetization 1358
 saturation of absorption 1358
 Sb–Te bond 1143
 scaling 1358
 scaling of integrated circuits 625
 scaling targets 627
 scanning electron microscopy (SEM) 11, 318, 343, 568
 scanning probe microscopy 1358
 scanning spreading resistance microscopy (SSRM) 414
 scanning thermal microscopy (SThM) 763
 scanning–tunnelling microscopy (STM) 660
 scattering
 – carrier–carrier 449
 – carrier–lattice 448
 – cross section 21
 – mechanism 420
 SCH laser 1358
 Schiff's base 940
 Schott glass code 76
 Schottky barrier 9, 701
 Schottky barrier diode SiC 547
 Schottky contact
 – capacitance-voltage characteristics 150
 – current–voltage characteristics 149
 – GaAs 524
 Schottky contacts 1358
 Schottky defect pairs 215

- Schottky diode
 - Ni-a-Si:H 591
- Schottky emission 707
- Schottky formation energy 216
- Schottky–Mott rule 147, 1358
- Schrödinger equation 1024
- Schroder van Laar equation 942
- schubweg 1123
- scintillator 250, 994
- scintillator material
 - growth 235
- screen 1358
- screen printing 1358
- screen printing (SP) 717, 1127
 - polymer thick films 724
- screen-printed paste 1099
- secondary electrons 1358
- secondary electrons (SE) 348
- secondary ion mass spectrometry (SIMS) 127, 311, 343
- secondary ion yields 381
- secondary-ion emission spectroscopy (SIMS) 846
- secondary-ion mass spectrometry (SIMS) 512, 1358
- secondary-ion mass spectroscopy (SIMS) 792
- Seebeck effect 730
- seed crystal 335
- seeded chemical vapor transport (SCVT) 335
- seeded physical vapor transport (SPVT) 335
- segregation 125, 264, 272
- segregation coefficient, k 1359
- segregation of components 674
- selection rule 1359
- selective
 - epitaxy 1359
 - etching 514
 - growth 493
- selenization 1359
- self-assembled monolayer (SAM) 1247
- self-assembly 1001, 1359
 - organic material 1231
- self-assembly method
 - inverse opal 1008
- self-defocusing 1359
- self-diffusion 130
 - dopants 128
- self-diffusion parameter 128
- self-focusing 1359
- self-interstitial configurations 103
- self-interstitials 128
 - concentration 122
 - semiconductor 122
- self-organized process 736
- self-scanned electronic readout 1122
- self-similar correlations 203
- Sellmeier equation 51
- semiconduction
 - CdTe film 703
 - filmelectronic properties 1108
 - layer 412
 - material 499
 - nanotubes 1151
 - thin film 659, 700
- semiconductor 3, 5, 13, 22, 51, 59, 288, 442, 725, 766, 805, 1000, 1084, 1149
 - absorber layer 1097
 - alloy 11, 338, 736, 741
 - alloy concept 736
 - amorphous 59
 - amorphous organic 1113
 - amorphous phase 1140
 - analysis 373
 - band gap 435, 1098
 - band gaps 8
 - band picture 1066
 - band structure 895
 - band-structure 152
 - binary compound 122
 - carrier scattering mechanisms 750
 - chalcogenides 1140
 - characterization 362, 373
 - characterization techniques 348
 - components 418
 - compound 8
 - conductivity 696
 - core levels 152
 - crystalline 38, 121
 - crystals 1066
 - cut-off wavelength 1097
 - density of states 163
 - detector 621, 1359
 - device 409, 538, 696, 888
 - device applications 3
 - die 1271
 - dielectric band gap 154
 - dielectric constant 154
 - diode lasers 895
 - direct band gap 890
 - dislocation 13
 - electrical characterization 410
 - electron transport 806
 - figures of merit 1081
 - film 678, 1244
 - first generation 537
 - free carriers 411
 - functional properties 343
 - future-generation 537
 - gain 907
 - gain/loss 892
 - gauges 1115
 - glass 162
 - Group III–V 131
 - Group IV 131
 - growth method 233
 - Hall effect 418
 - heterostructures 147, 154, 363, 365
 - ideal 154
 - impurities 63
 - indirect band gap 890
 - indirect-band-gap 442
 - industry 5, 255
 - interband transition region 749
 - interfaces 153
 - laser 12
 - material 891
 - materials 4
 - microelectronics 625
 - mobility 418
 - modulator 907
 - nanofabrication 1080
 - nondegenerate 22
 - nonlinear response 1083
 - n-type piezo-electric 336
 - optical amplifier (SOA) 889, 907
 - optical band gap 890
 - optical properties 1066
 - organic 1226
 - oxide sensing 1241
 - permittivity 416
 - physical properties 541
 - properties 577
 - pseudo-binary (CdTe–ZnTe) 338
 - quantum well 1021
 - quantum wells 907
 - refractive index 892, 1080
 - relative permittivity 22
 - resistivity 410, 411
 - response time 1083
 - silicon 234, 256
 - specific heat 764
 - spontaneous emission 890
 - spontaneous ordering 736
 - strain in 894
 - structural characterization 365
 - tetrahedral 895
 - thermal conductivity 393
 - thermal properties 741
 - traveling heater method 338

- undoped 890
- wafer 412
- work function 700
- wurtzite 740
- zincblende 740
- semiconductor alloy 9
- semiconductor laser 753, 888
 - current confinement 898
 - double heterostructure 897
 - edge emitting 896
 - index-guiding 898
 - radiative recombination 901
- semiconductor layer
 - analysis 376
 - H diffusion 128
 - hydrogen 127
 - rare-earth doping 273
 - sheet resistivity 411
- semiconductor material 411
- semiconductor nanocrystals 1359
- semiconductor third generation 538
- semiconductor–dielectric interface
 - electrical characterization 410
- semiconductor-grade silicon (SG-Si) 257
- semiconductor–insulator interfaces
 - charge pumping 428
- semiconductor–semiconductor interfaces 156
- semi-insulating 1359
- semi-insulating planar buried heterostructure (SIPBH) 899
- sense transistor 1185
- sensing
 - element 1228
 - filament 1255
 - material 1241
 - mechanism 1258
 - membrane 1250
 - membrane fabrication 1259
- sensitivity 1128, 1244
- sensitivity relative 382
- sensor 1131, 1155, 1241, 1249, 1257
 - amperometric 224
 - microelectronics 717
 - microstructure glass 1257
 - miniaturization 214
 - optical absorption 1251
 - organic vapour 1259
 - oxygen 223
 - solid-state 1244
 - thick film technology 728
- separate confinement heterojunction (SCH) 530
- shallow
 - acceptors SI GaAs 521
 - donor electron 783
 - impurities 445
 - impurity doping 448
 - structures 380
 - trap level 702
 - traps 1125
 - trench isolation (STI) 630
- shallow-energy-level dopant 1359
- shape
 - anisotropy 89
 - forming 611
- shear modulus 758, 1359
- sheet resistance 418, 1359
- sheet resistivity 411
- Sheik–Bahae relation 1066
- Shell Solar 1104
- shields 1359
- Shockley–Read 871
- short waveband (SW) 856
- short-range atomic structure 1359
- short-wavelength infrared (SWIR) 312
- shot noise 868, 1188
- showerhead reactor 285
- Si
 - conduction-band effective mass 445
 - diffusion coefficient of electrons 470
 - doping effect 450
 - drift velocity 469
 - electrical properties 451
 - Energy levels of impurities 444
 - equilibrium concentrations 108
 - Hall-effect parameter 459
 - interstitials 103
 - interstitial aggregation 113
 - interstitials 103, 108
 - nitrogen influence 113
 - optical absorption spectrum 477
 - oxidation kinetics 636
 - room-temperature resistivity 453
 - thyristors 551
- Si/SiGe quantum structures 488
- $\text{Si}_{1-x}\text{Ge}_x$ alloys 129
- $\text{Si}_{1-x}\text{Ge}_x$ system
 - electron transport properties 368
- Si/SiO₂ system
 - point defects system 639
- Si:H
 - conduction tail states 59
- Si-based diodes 539
- Si-based transistor 639
- Si-based transistor technology 635
- SiC 129
 - *a*-face growth 543
 - defects 543
 - device 542, 544
 - diffusion coefficient of dopant 544
 - electrical properties 543
 - electronic application due 542
 - high-frequency devices 544
 - IGBT (insulated gate bipolar transistor) 551
 - MOSFET 539, 550
 - packaging 544
 - polytypes 541
 - rectifiers 548
 - Schottky barrier 544
 - Schottky barrier diodes 547
 - stacking sequence 541
 - substrates 543
 - thyristors 551
- SiGaAs
 - electrical analysis of defects 518
 - MESFET 525
 - shallow-level defects 519
 - TSC spectrum 519
- SiGe 355
 - growth 492
 - hydrogen passivation 492
 - in-situ hydrogen bake 492
 - optical properties 488
 - pre-epitaxy cleaning 492
 - quantum wells 488
- signage 993
- signal decay rate 1359
- signal processing 1076
 - optical 1076
- signal processing device 1078
- signal switching 1076
- signal-to-noise ratio (SNR) 1133
- SiH₃
 - enhanced surface diffusion 588
- silane (SiH₄) 290, 581
- silica glass 569
 - PbO 1069
 - surface roughness 1045
- silicon 4, 5, 7, 913
 - amorphous 59
 - bandgap 255
 - crystal 256
 - crystals growth new methods 266
 - CZ crystal growth 260
 - Czochralski (CZ) grown 112
 - device 5

- device electrical performance 1268
- device power dissipation 1268
- dislocation-free 256
- dopants 265
- doping 260
- element 256
- FZ method 259
- growth 267
- hydrochlorination 257
- lattice constant 482
- magnetic-field-applied continuous CZ 267
- materials preparation 256
- metallurgical-grade 257
- microelectronics 1152
- multiplexer 876, 878
- nitride 704
- nitride films 679
- nitrogen concentration 112
- n-type 260
- optical constants 472
- optical properties 472
- physicochemical characteristics 263
- power devices 539
- p-type 260
- semiconductor industry 255
- silica 256
- single-crystal growth 258, 260
- stacking faults 112
- tetrachloride 257
- silicon carbide (SiC) 13, 275, 537
- carrier mobilities 540
- domain walls 543
- lattice match 541
- material properties 540
- micropipes 543
- preparation 542
- silicon crystal growth
 - magnetic-field-applied CZ 266
- silicon polymorphous 142
- silicon surface capacitance 425
- silicon technology 3, 539, 1244
- silicon tetrachloride (SiCl₄) 257
- silicon wafer technology 102
- silicon-based light emitter 488
- silicon-germanium (SiGe) 7
 - epitaxial layers 492
 - growth 492
 - physical properties 482
- silicon-on-insulator (SOI) 421
- silicon-oxynitride
 - SiON 641
- silk screen method 1276
- silk screen printing 719
- silver conductors
 - thick film 721
- silver/palladium conductors
 - thick film 721
- simply thermoluminescence (TL) 989
- SIMS (secondary ion mass spectrometry) 373
- single crystal diamond 552
- single crystal growth 261, 361
- single crystal Si
 - optical constants 472
- single crystal silicon 259
 - conversion efficiency 1099
 - CZ growth 261
- single crystal substrates 9
- single electron tunnelling 1199
- single flux quantum 1212
- single photon counting 991
- single quantum wells (SQW) 332
- single-crystal aluminium nitride 786
- single-electron density of states 1198
- single-electron transistor (SET) 1151
- single-junction cells 1100
- single-layer organic photoconductor 1359
- single-photon emission computed tomography (SPECT) 994
- single-wall carbon nanotube (SWCN) 1148, 1225
- sintered ceramics
 - aluminium nitride 786
- sintering 611
- Si–O bond stretching vibrations 450
- Si–O bonding network 638
- SiO₂ 8
 - physicochemical properties 638
 - reliability 641
- SiON film 642
- S–I–S junction 1199
- Si–SiO₂ interface 428
- SIT (static induction transistor) 549
- slab waveguide 1001
- slider 1359
- sliding 873
 - boat 273, 1359
 - boat LPE 310
 - boat method 274
 - boat process 873
- slope parameter 1359
- slush 1359
- small-angle X-ray scattering (SAXS) 568
- smart sensor 1257
- smart windows 226
- smectic liquid crystal 947
- smectic phase 919, 1359
- sodium nitrite (NaNO₂)
 - order–disorder ferroelectric 600
- soft lithographic method 1232
- soft magnetic materials 91
- soft underlayer (SUL) 1175, 1359
 - materials 1182
- solar cell 13, 532, 565, 581, 731, 1360
 - amorphous silicon (a-Si:H) 1100
 - a-Si:H-based 591
 - band gap 1034
 - crystalline silicon 1099
 - GaAs 1101
 - heterojunctions 1101
 - μ c-Si:H-based 593
 - photocurrent 1034
 - photovoltaic 1095
 - single-junction 1101
 - thin-film polycrystalline silicon 1099
- solar cell photovoltaic 1095
- solar electric modules 1099
- solar energy 1095
- solder
 - alloy 1277
 - alternatives 1281
 - conduction 1282
 - dipping 727
 - electronic interconnect 1274
 - flip chip 1277
 - interconnect 1273
 - interconnect reliability 1275
 - intergranular failure 1275
 - intermetallic compound 1275
 - microstructure 1274
 - paste 1276
 - Pb-free product 1277
 - Pb–Sn 1277
 - plating 1276
 - RF circuit 1277
 - wetting 1276
- solder joint
 - flip chip 1276
 - microstructure 1275
 - strain 1275
- soldering 727
- sol-gel techniques 1070
- solid electrolyte 223–225
- solid ionic conductors 221, 226

- solid oxide electrolytes 215
- solid oxide fuel cells (SOFC) 224, 226
- solid polymers 389
- solid silicon
 - impurity solubility 264
- solid state ionics 213
 - future trends 226
- solid state recrystallization (SSR) 245, 303, 305
- solid/liquid/gaseous phase equilibria 1360
- solids
 - electrical behavior 221
- solid-source MBE
 - effusion cell 295
- solid-state lightening 13
- solid-state polymerization 1086
- solidus 1360
- solid–vacuum interfaces 153
- solitonic propagation of pulses 1360
- sonoluminescence 984
- sound velocity 545
- space charge region 221, 1211
- space system 538
- space transformer 1268
- space-charge-limited conductivity 701
- spacing layer 1163
- spatial frequency 1360
- spatial resolutions 375
- specific heat 764, 1360
 - III–V binaries 741
- specific volume 401
- spectral broadening 1068
- spectroscopic ellipsometry 488
- spectroscopic response 191
- specularity parameter 698
- spherical aberration 353, 1360
- spin coating 1245
 - film thickness 1247
- spin injection 99
- spin magnetic energy 43
- spin manipulation 98
- spin polarization ratio 1186
- spin split-off band 901
- spin valve 1162, 1167, 1360
 - memory application 1187
 - resistance 1168
- spin valve read head 1163
- spin-based electronics 79
- spin–charge separation 1151
- spinels 92
- spin-on-glass 1099
- spin–orbit coupling 985
- spin–orbit scattering 1167
- spin–spin scattering 1167
- spintronic devices 80
- spintronics 98
- splay deformation 926
- splay elastic constant 927
- spontaneous emission (SE) 896
- spontaneous ordering 736
- spontaneous polarization 598, 1360
- spreading resistance technique 410, 412, 414
- SPRITE 1360
- SPRITE detector
 - carrier concentration 864
 - carrier velocity 866
 - conductivity modulation 864
 - power dissipation 866
 - scan velocity 866
- sputter depth profile 1360
- sputter-induced roughness 382
- sputtering 345, 378, 668, 1360
 - asymmetric-AC 675
 - chemical information 378
- sputtering radio-frequency (RF) 675
- sputtering reactive 581, 674
- sputtering yield 669, 671
- square-root DOS 62
- squeegee 1360
- SRAM 1360
- SSR 1360
- SSR crystals 306
- stability parameter
 - superparamagnetic 1175
- stabilized a-Se 1125
- stacked capacitors 649
- Staebler–Wronski effect 576
- stainless steel
 - substrate 720
- standing wave 897
- Stark effect
 - quadratic 1029
 - quantum-confined 1028
- static dielectric constant 457
- static disorder 62
- static random-access memory (SRAM) 647
- static SIMS (SSIMS) 377, 378
- steady state photoconductivity (SSPC) 138
- steady-state electron transport 811, 820
- steady-state plasma 582
- STEBIC 362, 368
- Stepanov technique 1360
- stimulated luminescence 983
- Stirling cryocoolers
 - long term reliability 860
- Stirling engine refrigerator 857
- stoichiometry 1360
- Stokes Shift 1360
- stop band 999
- storage capacitor
 - leakage of charge 648
- storage densities 1144
- storage time 1172
- strain 1360
 - axial 895
 - components 1117
 - compressive 1026
 - gauge 1114
 - hydrstatic 895
 - relaxation 332
 - tensile 1026
 - tuning 1013
- strained bond formation 588
- strained quantum well 1025, 1360
- strained-layer techniques 907
- strain-induced modulation 1117
- Stranski–Krastanov (SK) growth 332
- Stranski–Krastanov mode 298
- stress 1360
- strong anchoring 1360
- structural characterization
 - semiconductor heterostructures 365
- structural parameters
 - crystal density 737
 - lattice parameter 737
 - lattice-matching condition 737
 - molecular density 737
- structural relaxation 1360
- structure factor 349
- sub-100-nm (CMOS) technology 643
- sublimation epitaxy
 - growth rate 542
- sublimation growth 245
- sublimation travelling heater method (STHM) 246
- submicron machining 1002
- substitutional impurities 1361
- substrate 1361
 - Al₂O₃ ceramic 1279
 - basic types 665
 - dielectric constant 1278
 - embedding of components 1280
 - flip chip interconnect 1279
 - lattice-matched 873
 - LTCC 1279
 - material 1279

- non-lattice-matched 315
- organic 1278, 1279
- Poisson's ratios 1115
- substrate debiasing 879
- substrate material 1022, 1279
- underfill 1280
- subthreshold slope 427
- SUMIT (simultaneous unipolar multispectral integrated technology) 880
- super-bandgap 1065
- superconducting oxide 1194
- superconducting state
 - resistance 1195
- superconductivity 1151, 1193
 - BCS theory 1198
 - history 1194
 - mechanism 1198
 - MgB_2 1214
 - organic materials 1195
- superconductor 1202
 - electrodynamics 1197
 - high-temperature 1193
 - magnetic behaviour 1196
 - metallic 1194
 - phase diagram 1203
 - phase diagrams 1201
 - zero field-cooled 1196
- superconductors of type I 1361
- superconductors of type II 1361
- supercontinuum 1068
- supercooled liquid 399
- supercooling temperature 1361
- superlattice (SL) 845, 883, 1021, 1024, 1083, 1162, 1361
 - THz-frequency emission 1031
- avalanche photodiode (SL-APD) 1035
 - cell size 1031
 - effective band gap 1035
 - mathematical description 1026
 - structures 124
 - vertical transport 1031
- superlattice avalanche photodiode (SL-APD) 1361
- superparamagnetic effect 1171, 1173
- superparamagnetic limit 1156, 1187
- superparamagnetism 1361
- super-resolution near-field structure (super-RENS) 1139
- supersaturated solutions 310
- supersaturation 109, 335, 683, 1361
- superstrate 1361
- supramolecular assembly 1248
- supramolecule 1241
- surface
 - absorption 587
 - anisotropy 90
 - band bending 424
 - energy 95
 - kinetics 683
 - mobility 686
 - morphology 279
 - plasmon 1252
 - reaction concept 583
 - roughness 1045
 - scattering 30
 - source 664
 - topography 374
- surface acoustic wave (SAW) 729
 - device 608
- surface alignment 925
 - director orientation 932
 - energy 926
 - homeotropic 925
 - hybrid 939
 - pre-tilt 928, 937
 - strong anchoring 926
 - uniform planar 925
 - weak anchoring 926
- surface analysis
 - dopant concentrations 377
- surface chemical analysis 373
 - ISO TC 201 373
- surface emitting laser structures (VCSEL) 296
- surface mount components 727
- surface mount devices
 - component 724
 - thick film techniques 724
- surface passivation 1361
- surface-mount devices (SMDs) 1361
- susceptibility 64, 86, 209
 - frequency dependence 205
 - linear 195
 - nonlinear 1063
 - self-similar scaling 205
- susceptibility increment 198
 - dipole contribution 209
- susceptibility relative
 - frequency-dependent 191
- susceptor 1361
- SW 1361
- Swanepoel's method 71
- s-wave symmetry 1203
- switching
 - thermally driven 1172
- switching field 1185
- switching time 627
- SWNT (single wall nanotube)

- electrical transport 1150
- intrinsic superconductivity 1151
- ropes 1150
- SWNZ (single wall nanotube)
 - electrical transport 1150
- synthetic antiferromagnet 1361
- synthetic antiferromagnetic film 1166, 1185
- synthetic ferrimagnetic media (SFM) 1361
- synthetic ferromagnetic media (SFM) 1173
- system performance 1131
- system-on-a-chip (SOC) 255

T

- T nonlinear figure of merit 1361
- TA Instruments 405
- tail states 1361
- target material 673
- Tauc gap 1362
- Te concentration 846
- Teal-Little (TL) 451
 - method 259
- technology node 1362
 - cycle time 626
- telecommunication 888, 997, 1032
- television screens 983
- TEM
 - imaging 353
 - lenses 352
- temperature coefficient of refractive index (TCRI) 51
- temperature coefficient of resistance (TCR) 1362
- temperature coefficient of resistivity (TCR) 697
- temperature-modulated differential scanning calorimetry (MDSC)
 - thermal conductivity 403
- template
 - self-assembled 1005
- templating 1001
- tensile strain 1026, 1362
- ternaries
 - III–V 740
- ternary alloy 1022, 1165
 - mobility 750
 - optical band gap 791
- ternary and quaternary alloys 1362
- ternary InGaN 792
- ternary parameter 736
- ternary systems 305
- Tersoff potential function 104

- tertiarybutylarsine (TBA) 286
- tertiarybutylphosphine (TBP) 286
- tetracyanoquinodimethane (TCNQ) 1224, 1249
- tetraethoxysilane (TEOS) 651
- tetragonal ferroelectric phase 605
- tetrahydrofuran (THF) 614
- tetrathiafulvalene (TTF) 1224
- TFT
 - fabrication 1108
 - inverted staggered 1112
 - longitudinal 1117
- thallium bromide 1130
- thallium-based compounds 1202
- thermal analysis 385
- thermal budget 1362
- thermal coefficient of expansion (TCE) 720
- thermal conductivity 385, 539, 545, 1362
 - bulk AlN 763
 - ceramics 391
 - crystalline insulators 391
 - crystalline polymers 394
 - crystalline semiconductors 393
 - CVD diamond 555
 - glasses 391, 393
 - lattice 393
 - metal alloys 391
 - metals 391, 395
 - nanotube 1148
 - noncrystalline insulators 393
 - polymeric composite 395
 - polymers 391
 - substrate 720
- thermal current 1030
- thermal decomposition 678
- thermal diffusion current 870
- thermal expansion
 - principle 396
- thermal expansion coefficient (TEC) 385, 396, 545, 761, 1362
 - Group III–V 741
- thermal expansion matching 544
- thermal imaging 857, 879
 - photovoltaic HgCdTe 857
- thermal nonlinearities 1362
- thermal poling 1069
- thermal resistivity
 - Group III–V binaries 742
- thermally grown SiO₂ 706
- thermally stimulated current (TSC) spectroscopy 1362
- thermally stimulated luminescence (TSL) 989
- thermionic emission 149
- thermionic emission current 416
- thermistors 729, 1362
 - positive temperature coefficient (PTC) 730
- thermochromism 921
- thermocompression 726
- thermocompression bonding 726
- thermodynamic melting temperature 398
- thermoelectric cooler 879
- thermomechanical fatigue 1275, 1362
- thermophotovoltaic (TPV) 286, 1362
- thermophysical properties 103
- thermoplastic polymer
 - adhesive 1281
 - glass-transition temperature 1281
- thermoplastic polymer thick films 724
- thermosetting
 - material 1281
- thermosetting polymer thick films 724
- thermosonic bonding 726
- thermotropic 1362
- thick film 613, 698, 717, 720, 721, 1244, 1362
 - copper 722
 - definition 717
 - dielectric pastes 723
 - fabrication 717
 - photoconductor 730
 - piezoelectrics 729
 - platinum conductor 728
 - polymer 723
 - substrates 720
- thick film chemical sensor 730
- thick film components
 - resistors 725
 - tolerance 725
- thick film conductor
 - characteristics 721
 - sheet resistivity 721
- thick film definition 717
- thick film fabrication 717
- thick film hybrid circuits
 - active components 725
 - manufacturing 717
 - packaging 727
- thick film platinum 722
- thick film resistor
 - material 722
 - pastes 722
- thick film solar cell 731
- thick film technology 718, 724
 - printing 718
- thick heteroepitaxial layers 782
- thin dielectric film 700
- thin film 47, 659, 696
 - cobalt alloys 1169
 - deposition 661
 - deposition methods 660
 - electrical properties 660, 696
 - fundamental properties 692
 - morphological characteristic 689
 - optical properties 660, 692
 - optics 692
 - organic 1233, 1241, 1248
 - properties 660, 682
 - resistivity 31
 - semiconductor 1107
 - solar cells 1103
 - stack transmission properties 695
 - structure 682, 683
 - technology 659
 - transducer 1114
 - transistor (TFT) 565, 581, 590, 660, 1107, 1121
- thin film deposition
 - variable 682
- thin film phthalocyanine 1254
- thin metallic films 696, 698
- thin organic film 1245
 - island density 1245
- third generation semiconductor 538
- third harmonic generation 1069
- third-order nonlinearities 1070
- third-order nonlinearity
 - organic molecules 1086
- Thomas–Fermi screening 1211
- threading dislocations 763, 1362
- threading screw dislocations 543
- three-dimensional lattice 386
- three-terminal structure 1226
- three-valley Monte Carlo simulation 806
- three-zone model
 - microstructure 690
- threshold 898
- threshold current 903
 - characteristic temperature 904
- threshold electric field 928
- threshold voltage (V_{th}) 427, 1112, 1117, 1362
 - modulation 1118
- thyristors 539
- Ti/SiC Schottky diode 548
- tight-binding model 1026

- tilted perpendicular recording
 - geometry 1181
- time delay and integration (TDI) 867
- time division multiplexing 1362
- time-dependent dielectric breakdown (TDDb) 640
- time-domain charge measurement (TDCM) 519, 1362
- time-of-flight (TOF) 144, 572, 959
 - electron mobility 176
 - post-transit current 145
- time-of-flight SIMS (ToFSIMS) 377
- tipping 873, 1362
 - boat 273
- tipping techniques 308
- toner 955
- total heat flow 404
- track density 1174, 1184
- track width 1170
- transconductance 427, 640
- transfer mold process 1280
- transfer reaction 679
- transferred electron effect 1362
- transformer 92
- transient electron transport 810, 824
 - characteristics 821
 - drift velocity 821
- transient enhanced diffusion (TED) 124
- transient photoconductivity (TPC) 142, 143
- transistor 3–5, 410, 725, 1363
 - average price 7
 - in-plane 553
 - performance 9
 - p-n junction 9
 - scaling 628
- transistor channel scaling 630
- transistor design technology
 - switching time 627
- transistor electrical characteristics 626
- transistor gate dielectric
 - industry scaling 632
- transistor performance 632
- transistors on a single chip 7
- transit time 1363
- transition
 - dipoles 191
 - insulator–metal 445
- transition probabilities
 - relative 986
- transition region entropy 201
- transition shape 1180
- transition temperature 920, 1193
 - liquid crystal–isotropic 924
 - nematic–isotropic 920
- transition-metal films 1162
- transition-metal ions 82, 986
- transition-metals
 - magnetic properties 97
- transmission coefficient 70, 1186
- transmission electron diffraction (TED) 351
- transmission electron microscope (TEM) 343, 568, 588
- transmission electron microscopy (TEM) 11, 689, 846
- transmission line measurement (TLM) 410
- transmission line model (TLM) 417
 - test structure 417
- transmission modulator 1036
- transmission probability 1030
- transmittance 71
- transparent conducting oxide (TCO) 1102, 1363
- transparent films 695
- transparent layer 694
- transport
 - agent 333
 - bipolar 1030
 - chamber 317
 - currents 1211
 - energy 174, 182
 - unipolar 1030
 - vertical 1030
- transport coefficients 170, 179
- field dependence 180
- transport phenomena
 - low-dimensional 1150
- transport-limited growth 286
- transverse electric (TE) gain 896
- transverse magnetic (TM) gain 896
- transverse magnetic bias film 1363
- transverse optic (TO) 55, 759
 - modes 741
- trap concentration 703
- trap distribution 1113
- trap level 702
- TRAPATT diode 471
- trap-free space-charge-limited current density (TFSLC) 962
- trap-free transport 962
- trapped carrier 1135
 - density 1110
- trapping 871
 - levels 143
- trapping centers 1363
- traveling heater method (THM) 237, 245, 303, 306, 1363
 - ion implantation 866
- triangular lattice 1000
- triboluminescence 984
- trichlorosilane (SiHCl_3) 257
- triethylamine (TEN) 289
- triethylgallium (TEGa) 287
- trimethyl-aluminium (TMA) 508
- trimethylantimony (TMSb) 287
- trimethyl-gallium (TMG) 508
- trimethyl-indium (TMI) 508
- trimming 1363
- triode sputtering 673
- triphenylenes 924
- triple-junction cell 1101
- triple-layer graded heterojunction (TLHJ) 312
- trisazo compound 972
- true specific heat capacity 390
- tungsten bronze
 - PbNb_2O_6 (lead metaniobate) 605
- tuning photonic crystal response 1013
- tunnel conductivity 1203
- tunneling conduction 1363
- tunneling current 634
- tunneling magnetoresistance (TMR) 1188, 1363
- tunneling transition
 - charge carrier 179
 - coefficient 1186
 - probability 169
- tunnelling single electron 1199
- turbo-disc reactor 285
- twin crystal 1363
- twinning 315
- twist deformation 926, 927
- twisted nematic display
 - chiral dopant 936
- two-beam experiment 1067
- two-dimensional electron gas (2DEG) 297
- two-photon absorption 1067, 1079, 1084, 1363
- two-zone furnace 246
- type of film 682
- type I band alignment 1024
- type I low- T_c superconductor 1196
- type II band alignment 1024
- type II superconductor 1200
 - mixed state 1201
 - pinning 1212
 - vortex–vortex interaction 1211
- Tzero technology 405

U

UHV 1363
 ultrafast nonlinear response 1363
 ultrafast nonlinearity real part 1084
 ultrahigh recording densities 1181
 ultrahigh-vacuum chemical vapour deposition (UHV-CVD) 481
 ultrahigh-vacuum electron-cyclotron resonance-radio-frequency magnetron sputtering (UHVECR-RMS) 774
 ultra-large-scale integration (ULSI) 255
 ultrasonic 726
 ultrasonic bonding 726
 ultrathin films 1212
 ultraviolet light absorption spectroscopy (UVLAS) 582
 Umklapp process 392
 unbalanced magnetron technique 677
 under-bump metallurgy (UBM) 1276, 1278
 underfill 1363
 – material 1280
 undoped GaAs 510
 uniaxial anisotropy 1363
 uniform planar alignment 1363
 unipolar avalanche photodiode (UAPD) 1035, 1363
 unit cell 682
 upper critical field (B_{c2}) 1363
 Urbach tail 62
 UV lithography 1002
 UV poling 1069
 UV sensors 754

V

V/I boundary 1363
 vacancy 1363
 – aggregation 109
 – concentration 110, 115, 123, 125
 – diffusivity 112
 – flux 124
 – formation energy 104
 vacuum deposition 663
 vacuum distillation 838
 vacuum field effect transistor (VFET)
 – diamond 554
 vacuum permittivity 48
 valence band 1363
 valence band (VB) 57, 890, 1221
 – amorphous semiconductors 566

 – deformation potentials 747
 – offset 152
 valence band offset 154, 157
 valence electrons 60
 valence-alternation pairs (VAP) 567
 van der Pauw technique 413
 van Hove singularities 59
 vapor growth 246
 vapor growth techniques 1363
 vapor pressure controlled Czochralski (VCZ) 244
 vapor sensing 1254
 vapor-phase epitaxy (VPE) 6, 325, 508, 873, 889, 1364
 variable-range hopping (VRH) 172, 705, 1113
 (VCSEL) 1364
 vector flow epitaxy (VFE) 285
 Vegard's law 1364
 vehicle 1364
 velocity–field characteristic 811
 – AlN 813
 – GaN 813
 – Group III–V nitride semiconductor 813
 – InN 813
 Verdet constant 69
 Verneuil growth method 233, 237, 250
 Verneuil technique 1364
 vertical dipper LPE 876
 vertical gradient freeze (VGF) 237, 245, 506
 – GaAs 506
 vertical transport 1030, 1364
 vertical unseeded vapor growth (VUVG) 246
 vertical-cavity surface-emitting laser (VCSEL) 531, 889, 1021, 1032, 1364
 – active region 905
 – beam profile 905
 – semiconductor–air interface 905
 vertical-gradient freeze technique 1364
 VI recombination 1364
 via-hole device 875
 – Group (I–V) characteristics 869
 vibration dynamics 207
 vibrational mean energy 386
 virtual gap states 1364
 virtual gap states (ViGS) 153
 viscosity coefficient
 – Miesowicz experiments 930
 – nematodynamics 931
 viscous torque 931

void 1364
 – density 111
 – size 107
 Volmer–Weber (VW) growth 332
 voltage noise measurement 432, 433
 – low-impedance devices 433
 voltage scaling 629
 vortex structures 1187
 VPE 1364
 VPE growth
 – group II–VI wide bandgap 330
 – source materials 330

W

W nonlinear figure of merit 1364
 wafer fusion 1008
 Wannier–Mott excitons 63
 waste in electrical and electronic equipment (WEEE) 1277
 wave
 – in-plane 1027
 wave function
 – Bloch-type 442
 wave soldering 727
 wavelength dispersive X-ray analysis (WDX) 343
 wavelength division multiplexing 1364
 weak absorption tail (WAT) 62
 weak links 1364
 web photoreceptors 1364
 white-light LED 13, 992
 wide bandgap
 – compounds 325, 332
 – electron mobilities 774
 – epitaxial growth 328
 – Group II–VI compound semiconductors 327
 – Group II–VI compounds 328
 – InN 774
 – materials 781
 – quantum dot 332
 – quantum well 332
 – semiconductor 147, 158, 537, 538
 wide-band X-ray 839
 Wiedemann–Franz–Lorenz law 395
 windows growth 235
 wire bonding 726, 1364
 – digital applications 1272
 – materials 1272
 work function 700, 1030, 1364
 write efficiency 1364
 write gap 1364

write head 1364
 – efficiency 1161
 write pole tip 1180
 wrong bond 1364
 wurtzite
 – AlN 809
 – GaN 761, 811
 – Group III–V binaries 747, 748
 – Group III–V semiconductors 747
 – structures 755
 Würtzite structure 1364
 WZ-GaN
 – specific heat 765

X

xerographic discharge 965, 1364
 xerographic gain 1365
 xerography 955, 1365
 XeroX Copier 955
 X-ray 11, 1365
 X-ray absorption fine-structure (XAFS) spectroscopy 1141
 X-ray absorption near-edge structure (XANES) 1141
 X-ray absorption spectroscopy (XAS) 758
 X-ray damage 1124
 X-ray diffraction 1365
 X-ray diffraction (XRD) 331, 354, 523, 1141
 – crystal plane spacings 350
 X-ray emission 344
 X-ray fatigue 1124

X-ray image
 – clinical 1130
 – detector 1123
 – photoconductor 1130
 – sensors 1124
 X-ray lithography 1007
 X-ray photoconductor 1122, 1123
 – absorbed energy 1131
 – material properties 1131
 – properties 1123
 X-ray photoelectron spectrometry (XPS) 343
 X-ray photoelectron spectroscopy (XPS) 373, 636, 1365
 – primary excitation 375
 X-ray sensitivity 1365
 X-ray storage 983
 – phosphor (XRSP) 993
 X-ray topography 1365
 X-ray-sensitive electron-beam image tube (XEBIT) 1130

Y

YAG (yttrium aluminium garnet) 726
 yellow luminescence 785
 yield strength 1365
 Young's modulus 757, 1365
 – semiconductors 740
 yttrium stabilized zirconia (YSZ) 547
 Y-type construction 1246
 Y-type deposition 1231

Z

ZBLAN glass 389
 Zeeman energy 81, 88
 zero-phonon transition 988
 zinc compound 684
 zinc selenide (ZnSe) 12
 zincblende structure 500, 755, 1365
 zirconia auto exhaust sensor 223
 zirconia based systems 225
 Zn in GaAs
 – diffusion profiles 512
 ZnBeSe
 – Be content 848
 ZnO
 – deep green luminescence 849
 – MBE-grown 850
 – optical properties 850
 – p-type doping 849
 – quantum wires 851
 – single crystals 336
 – undoped 850
 – wide band gap 843
 ZnSe 849
 – delta-doping 843
 – DX centers 843
 – epitaxial layer 330
 – N doping 843
 – p-type doping 843
 – single crystal 338
 ZnSe-based devices 843
 zone refining 1365
 zone-refining 5